

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 09:48:15 ; Search time 20 seconds  
(without alignments)  
2193.167 Million cell updates/sec

Title: US-09-905-589A-2  
Perfect score: 2364  
Sequence: 1 MRKISNRGSLRVAKVAYPLG.....ALGAIFHYIDSLNRQKSPAS 456

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:\*  
1: Pirl:\*  
2: Pirl2:\*  
3: Pirl3:\*  
4: Pirl4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	590.5	25.0	479	T23508	hypothetical prote
2	508	21.5	568	T32109	probable guanine
3	498	21.1	518	A40732	guanosine-diphosph
4	485	20.5	454	JC4616	aprase (EC 3.6.1.
5	442.5	18.7	455	S48859	nucleoside triphos
6	372	15.7	516	G84442	probable nucleosid
7	343	14.5	485	T34147	hypothetical prote
8	339.5	14.4	483	D8276	hypothetical prote
9	329.5	13.9	630	S50463	hypothetical prote
10	307	13.0	557	T16696	hypothetical prote
11	301	12.7	405	E8276	hypothetical prote
12	285.5	12.1	510	I56242	lymphoid cell acti
13	281.5	11.9	572	T40856	probable nucleotid
14	245.5	10.4	1052	T04439	hypothetical prote
15	219	9.3	508	C86276	7A.9.33 protein -
16	140.5	5.9	628	A55421	nucleoside-triphos
17	111	4.7	679	C71007	probable formate d
18	108.5	4.6	1186	T12737	tail protein - Met
19	106.5	4.5	665	E75124	formate dehydrogen
20	104.5	4.4	322	T24948	hypothetical prote
21	103.5	4.4	322	T27308	hypothetical prote
22	102.5	4.3	440	I39847	cell-division prot
23	102	4.3	475	G72274	glutamyl tRNA-Gln
24	102	4.3	623	G81346	heat shock protein
25	102	4.3	4162	T42633	connectin/titin -
26	100	4.2	1233	T37045	nitrate reductase
27	100	4.2	26926	I38344	titin, cardiac mus
28	99.5	4.2	470	I31839	dihydrolipoamide d
29	99.5	4.2	504	S17724	H+-transporting tw

30 99.5 4.2 633 2 I39585  
31 99.5 4.2 2411 2 A46299  
32 98.5 4.2 633 2 AE2591  
33 98.5 4.2 633 2 E97373  
34 97.5 4.1 415 1 C64473  
35 97.5 4.1 548 2 S32615  
36 97 4.1 2611 2 T14591  
37 96.5 4.1 515 2 T35970  
38 96 4.1 711 2 S73898  
39 96 4.1 1217 2 T25894  
40 95.5 4.0 325 2 T32940  
41 95.5 4.0 325 2 D87733  
42 95 4.0 527 1 S25478  
43 95 4.0 549 2 S23448  
44 95 4.0 15281 2 S41309  
45 94.5 4.0 489 2 F97119

ALIGNMENTS

RESULT 1  
T23508  
hypothetical protein K08H10.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 17-Mar-2000  
C:Accession: T23508  
R:Gardner, A.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19750  
A:Accession: T23508  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-479 <WIL>  
A:Cross-references: EMBL:Z83113; PIDN:CAB05544.1; GSPDB:GN00023; CESP:K08H10.4  
C:Genetics:  
A:Gene: CESP:K08H10.4  
A:Map position: 5  
A:Introns: 36/3; 83/3; 189/1; 300/2; 412/3  
C:Superfamily: nucleoside triphosphatase chromatin-associated

Query Match 25.0%; Score 590.5; DB 2; Length 479;  
Best Local Similarity 34.2%; Pred. No. 2.1e-40;  
Matches 134; Conservative 78; Mismatches 155; Indels 25; Gaps 9;  
Qy 73 FYGIMEDAGSTGRVHVFOFTRPRETP-----TLTHETFKAVKPGISAYADDDVERSAQG 127  
Db 42 FFIIVIDAGSTGRHLHYKFIHDPATASHGMPKVEKEIFQEVKPGLSSEFAKSPSSAADS 101  
Qy 128 IRELLDVAQDIPDFFWKATPLVLKATAGURLLPGEKAQKLQKVE-VFKASPFVLVGGD 186  
Db 102 LEPLQARKEVPHFMWEKTPITLKATAGURLLPGNMADDILESVEERIFNSOFFFAAFPD 161  
Qy 187 CVSTMNCTDGVSAWITINFLTGLS-----KTPGGSSVGMLDLGGGSGTQIAFLPRV 237  
Db 162 AVNMPGSGVSWFTLNILLETFLFDETVGKPAHRSVAADF.LGGSGTQLTYPNN 221  
Qy 238 EGTLLQASPPGYLTALRMFNRTYKLYSYSLGLGLMSARLAILGVEGQPAKQKGLVSPC 297  
Db 222 EAVP-SEHVGVERDIDFFGHIRLFTSHFLGNGLIAARLNILQLETDNE-ESTHQLTSC 280  
Qy 298 LSPSEK-GEVHEAVTVRVSGKAAASLHELCAARVSEVLQ-NRVHRTVEVKHVDVAFS 355  
Db 281 MPEGYQLTEWEYALKFWNING--SSSHSFESCVGTTKNFVSEIEMRLRELKSGPVLFS 338  
Qy 356 YYYDLAAGVGLIDAEKGGSLVWGDFFEAARKVCTLTETQFQSS----PFSCMDLTYY-SL 410  
Db 339 YFFPRALNSGLVKNEGKIELRQFKEAETACRRETEIDDGSHWMPWQCLDLYIVSL 398  
Qy 411 LLOFPGPRSKVLKLRKIDNVETSWALGAF 442  
Db 399 LRDGVQFEDNQPLVLAKKIKGMEVSWGQGLAF 430



C:Accession: JC4616; PC4147  
R:Handa, M.; Guidotti, G.  
Biochem. Biophys. Res. Commun. 218, 916-923, 1996  
A:Title: Purification and cloning of a soluble ATP-diphosphohydrolase (Apyrase) from potato  
A:Reference number: JC4616; MUID:96158985; PMID:8579614  
A:Accession: JC4616  
A:Molecule type: mRNA  
A:Residues: 1-454 <HAN>  
A:Cross-references: GB:U58597; NID:g1381632; PIDN:AAB02720.1; PID:g1381633  
A:Accession: PC4147  
A:Molecule type: protein  
A:Residues: 59-95;96-131;132-160;236-253;332-345 <HA2>  
A:Experimental source: tubers  
A:Note: The authors translated the codon GCA for residue 215 as Gly  
C:Comment: This enzyme belongs to a family of E-type Apyrases, and it catalyzes the hydrolysis of ATP to ADP and inorganic phosphate. It has a high affinity for ATP and is insensitive to vanadate. The enzyme has nucleotide substrate specificity, divalent cation requirement, and is involved in starch synthesis.  
C:Genetics:  
C:Gene: rrop1  
C:Superfamily: nucleoside triphosphatase chromatin-associated  
C:Keywords: glycoprotein; hydrolase; transmembrane protein  
F:1-30/Domain: signal sequence #status predicted <SIG>  
F:8-25/Domain: transmembrane #status predicted <TM>  
F:31-454/Product: ATP-diphosphohydrolase #status predicted <MAT>  
F:44-65/Region: actin-heat shock protein 70-hexokinase beta-phosphate binding  
F:44-65/Region: nucleotide binding #status predicted  
F:192-212/Region: actin-heat shock protein 70-hexokinase gamma-phosphate binding  
F:192-212/Region: nucleotide binding #status predicted  
F:390-410,427-446/Region: hydrophobic carboxyl end  
F:151,262/Binding site: carbohydrate (Aen) (covalent) #status predicted

Query Match 20.5%; Score 485; DB 2; Length 454;  
Best Local Similarity 31.9%; Pred. No. 8.8e-32;  
Matches 138; Conservative 68; Mismatches 183; Indels 44; Gaps 12;

Qy 59 AHSPLGTAADGHE-VFYGMFMAGSTGTRVHFQFTRPPREPTLTH-ETFKAVKPGLSA 116  
Db 30 AQPLRRHLLSHSEHYAVIFDAGTSGSRVHVFDEKLGILPIGNNIEYFATEPGLSS 89  
Qy 117 YADDVEKSAQGTRELLDVAKQDIPDFWKATPLVLKATAGLLPGEKAQKLQKVEVF 176  
Db 90 YAEADPKAAANSLFPLLDGAEVVPQEQSETPLELGATAGLRLKGDAAEKLQAVRNIV 149  
Qy 177 K-ASPLVGDGCVSMNGTDEGVSAMITNFTLSKTPGSSVGMGLDGGGSGTOIAFLP 235  
Db 150 KNOSTFHSKDDQWITLDGQESYMAALNLLGNLKGDKYSTATTIDLGSSGVQAVAI 209  
Qy 236 RVEGTLLQA----SPPGYLTALRMFNRTYKLYSYSLGLGMSARLAILGGVGEQPAKDGK 291  
Db 210 SNEQFAKAPQNEDEGPVQKHLMSKDYLYVHSYLYNGQLAGRAEIF-----KASR 261  
Qy 292 ELVSPCLSPSPFKGEWEHAETVYRVSQKAAASLHELCAARVSEVLO-NEVHRTSEVK--- 347  
Db 262 NESNPCALEGCDGYSYGGVDYKVPKPKGSS-WKRCRLTRHALKINAKCNIECTFNG 320  
Qy 348 -----HVDYFAPSYYYDLAAGVGLIDAEGKSLVWG-DFEIAAKVVCRT----- 390  
Db 321 VNWGGGGDGGKXNHASFFVIDGAQGVIVDTKFPSSALAKPIQYLNAAKACQTNVADIKS 380  
Qy 391 --LETQPSQSPFCMDLTY-VSLLQEFPGPRSKVLKTRKIDN-----VETSWALGAIFH 443  
Db 381 IFPKTQDRNPYLCDMLIYEYTLVDGFLGNPKHEITVHDVQYKNLYVGAANPLGCAID 440  
Qy 444 YIDSLNFKQSPAS 456  
Db 441 LVSSSTTKIRVAS 453

RESULT 5  
S48859  
nucleoside triphosphatase precursor, chromatin-associated - garden pea  
C:Species: Pisum sativum (garden pea)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 29-Sep-1999

C:Accession: S65147; S48859  
R:Heieh, H.L.; Tong, C.G.; Thomas, C.; Roux, S.J.  
Plant Mol. Biol. 30, 135-147, 1996  
A:Title: Light-modulated abundance of an mRNA encoding a calmodulin-regulated, chromatin-associated triphosphatase  
A:Reference number: S65141; MUID:96197404; PMID:8616230  
A:Accession: S65147  
A:Molecule type: mRNA  
A:Residues: 1-455 <HS2>  
A:Cross-references: EMBL:Z32743; NID:g563611; PIDN:CAA83655.1; PID:g563612  
C:Superfamily: nucleoside triphosphatase chromatin-associated  
C:Keywords: nucleus

Query Match 18.7%; Score 442.5; DB 2; Length 455;  
Best Local Similarity 29.2%; Pred. No. 2.7e-28;  
Matches 125; Conservative 79; Mismatches 145; Indels 79; Gaps 15;

Qy 74 YGIMFAGSTGTRVHFQFTRPPREPTLTH-----ETFKAVKPGLSAYADDYKSAQGI 128  
Db 44 YAVFDAGSTGSRHVVHF-----NQNLDLLHIGKGVYNNKITPGLSSVANNPEQAASL 99  
Qy 129 REILLVAKQDIPDFWKATPLVLKATAGLLPGEKAQKLQKVEVF-KASPLVGDCC 187  
Db 100 IPLLQAEDEVVDDLOPKTEPRLGATAGLLNGDASEKILQSVDRMLSNRSTFNVPQDA 159  
Qy 188 VSMNGTDEGVSAMITNFTLSKTPGSSVGMGLDGGGSGTOIAF-----LPRVE 238  
Db 160 VSIIDGQESYLYWTVNYALGNLKKYTYGVIVDLGGSSVQMAVAVSKTKAKNPKV- 218  
Qy 239 GTLQASPPGYLTALRMFNRTYKLYSYSLGLGMSARLAILGGVGEQPAKDGKELVSPCL 298  
Db 219 --ADGDDP-VIKKVLKGIPIYDLYVHSYLYHFGREASRAEILKLTFRSP-----NPCL 267  
Qy 299 SPSPKGEWEHAETVYRVSQKAAASLHELCAARVSEVLONRVHRTSEVKH----- 348  
Db 268 LAGFNGIYTSGBEFKATATYSGANFNK-----CKTIRKALKLNYPQYQNTFG 318  
Qy 349 -----VDFYAFSYYYDLAAGVGLIDAEGKSLV-VGDFFIAAKVVCRT-TLETQP 395  
Db 319 GIWNGGGGQKQLFASSTFFPLPEDTGMVDASTPNEFILRPVDIETKAEACALNPFEDAK 378  
Qy 396 QSSPF-----SCMDLTVSLLLOEFGFPSPKVLKLT--RKIDN-----VETSWALGA 440  
Db 379 STYFPLDKKNVASYVCMDLIYQVYLLVD-GFGLDPLQKITSGKEIEYQDAIVEANWPLGN 437  
Qy 441 IFHYDLSL 448  
Db 438 AVEAISAL 445

## RESULT 6

G84442

probable nucleoside triphosphatase [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 18-Nov-2002

C:Accession: G84442

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: G84442

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-516 <STO>

A:Cross-references: GB:A5002093; NID:g3461821; PIDN:AAC32915.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2G02970

A:Map position: 2

C:Superfamily: nucleoside triphosphatase chromatin-associated

Query Match 15.7%; Score 372; DB 2; Length 516;

Best Local Similarity 29.0%; Pred. No. 1.9e-22;

Matches 113; Conservative 67; Mismatches 173; Indels 36; Gaps 13;

QY 74 YGIMPDAGSTGRVAVPOFT---RPPREPTLTHETFKAVKPGLSAVADVKEKSAQGR 129  
 Db 68 YSVVIDGSGTGRVAVPOFT---RPPREPTLTHETFKAVKPGLSAVADVKEKSAQGR 126  
 QY 130 EILDVAKQDIPDFWKATPLVLKATAGRLLPGEKQKLLQKVEFKASPFVGDGCS 189  
 Db 127 ELVEFAKGRVPGKMTTEVRLMATAGRLLEPVOEKILGVARRVLLKSGFLFEDEMAS 186  
 QY 190 INMGDEGVSAMITINFLTSGLSKTPGSSVGMULDGGSGTQIAPLPRVETGLQASDPGYL 249  
 Db 187 VISGSDGVAVMAYANPALGSLGDDPLKTTGIVELGASAQVTFV-----SEPEMPPEPS 241  
 QY 250 TALRMFRTYKLYSYSLGSLMSARLAIIGVEG-----QPADGKEIVSPCLSPSP 302  
 Db 242 RTISFGVNTVLYSHSFHFGQNAADKMLGSLSDHNSAVPEPTR-KITDPGAPKGY 300  
 QY 303 K---GEMEH-----AEVTVRSGQKAASLHELCAARVSEVLQNRVARETEVGVDFYAF 354  
 Db 301 NLDANTQKHLGSLAELS-RLSDSFGKGNYSQCRSAALTLIDGNGRI---LIIAGP 355  
 QY 355 SY-YTDLAAGVGLDAEKGSLLVGDPELIAKYVCTLETPQSPSCMDLTY-VSLLI 412  
 Db 356 SFLFFGLGKRWALSNMISAGBFCG--EDWSKLRVMDPSLHEBDLRLRYCFSSAYIVSLH 413  
 QY 413 QEGFEP-RSKVLLKTRKIDNVETSMALGA 440  
 Db 414 DTLCGLPDDERIKYANAGDIPDMALGA 442

RESULT 7  
 T34147  
 hypothetical protein C33H5.14 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 17-Mar-2000  
 C:Accession: T34147  
 R:Bradshaw, H.; Stelleyes, L.  
 submitted to the EMBL Data Library, November 1995  
 A:Description: The sequence of C. elegans cosmid C33H5.  
 A:Reference number: Z21482  
 A:Accession: T34147  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-485 <BRA>  
 A:Cross-references: EMBL:U41007; PIDN:AAA82272.1; CESP:C33H5.14  
 A:Experimental source: strain Bristol N2  
 C:Genetics:  
 A:Gene: CESP:C33H5.14  
 A:Introns: 22/1; 83/1; 120/2; 167/1; 269/3; 399/3  
 C:Superfamily: nucleoside triphosphatase chromatin-associated

Query Match 14.5%; Score 343; DB 2; Length 485;  
 Best Local Similarity 25.8%; Pred. No. 4.2e-20;  
 Matches 113; Conservative 74; Mismatches 157; Indels 94; Gaps 16;

QY 70 HEVFGIMPDAGSTGRVAVPOFT---RPPREPTLTHET---FKAVKPGLSAVADV 121  
 Db 21 NNKYGVCIDAGSSGTR--LFTVYTLKPLSGGLTIDTLIHSEPVVKKVPPGLSSPDKP 78  
 QY 122 EKSAGIRELDVAKQDIPDFWKATPLVLKATAGRLLPGEKQKLLQKVEFKASPF 181  
 Db 79 EGVAVYLPRLRFAEHIPEGLGTDLLIFATAGMRLLPKADAIITKLNQNLKSVTA 138  
 QY 182 L-VGDDCVSYNMGTDGVSAMITINFLTSGLSKTPGSSVGMULDGGSGTQIAPLPRVET 240  
 Db 139 LRVSSNRIIDGAEIGYSWIAVNYIIGRPDKENDSVYGMIDMGGSVQIAF--ETANE 196  
 QY 241 LQASPGVYLTALRM-----FNRTKLYSYSLGSLMSARLAIIGVEGQPADGKEIV 294  
 Db 197 KESYNGAVYEINLSIENEDYKTKITSTPLGSGAEG---LKKYENLVVSGNS-- 250  
 QY 295 SFKGEWHEAVTVRVSGQKAASLHELCAARVSEV----- 335

Db 251 NDSCSPGRLRLIEFTYNGEY-----DVCLAQVSSLGDKAQP 291  
 QY 336 -----LQNRVARETEVGVDFYAFSYDDLAAVGLIDAEKGSLLVGDPELIAKYV 387  
 Db 292 SCNPPTCELRNVIAPSVSLSTVQLYGFSEVWYTTSNFG-----SGGEYHQKFTDEVKRY 346  
 QY 388 CR-----TLETPQSPF-----SCMDLTVSLLQFSGPGRSKVLTTR---KID 430  
 Db 347 CQKMDIQQGFKKXEPFNADIERLGNCFKAANVTSLVD--GVNDKTKHLPQSVLXIA 405  
 QY 431 NVETSMALGALFHYIDSL 448  
 Db 406 GEEMQWALGAMLYHSKDL 423

RESULT 8  
 D86276  
 hypothetical protein F7A19.34 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 18-Nov-2002  
 C:Accession: D86276  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federapfel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chinn, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hulzar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzilli, R.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, K.; Xu, D.; Xu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: D86276  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-483 <STO>  
 A:Cross-references: GB:AE005172; NID:95080801; PIDN:AAJ39311.1; GSPDB:GN00141  
 C:Genetics:  
 A:Map position: 1  
 C:Superfamily: nucleoside triphosphatase chromatin-associated

Query Match 14.4%; Score 339.5; DB 2; Length 483;  
 Best Local Similarity 26.5%; Pred. No. 8.1e-20;  
 Matches 108; Conservative 78; Mismatches 166; Indels 55; Gaps 13;

QY 74 YGIMPDAGSTGRVAVPOFT---RPPREPTLTHETFKAVKPGLSAVADVKEKSAQGR 129  
 Db 70 YSVLIDAGSGTGRVAVPOFT---RPPREPTLTHETFKAVKPGLSAVADVKEKSAQGR 128  
 QY 130 EILDVAKQDIPDFWKATPLVLKATAGRLLPGEKQKLLQKVEFKASPFVGDGCS 189  
 Db 129 KLYERAKQDIPDFWKATPLVLKATAGRLLPGEKQKLLQKVEFKASPFVGDGCS 188  
 QY 190 INMGDEGVSAMITINFLTSGLSKTPGSSVGMULDGGSGTQIAPLPRVETGLQASDPGYL 249  
 Db 189 VISGSDGVAVMAYANPALGSLGDDPLKTTGIVELGASAQVTFV-----SEPEMPPEPS 243  
 QY 250 TALRMFRTYKLYSYSLGSLMSARLAIIGVEGQPADGKEIVSPCLSPSP 302  
 Db 244 RTISFGVNTVLYSHSFHFGQNAADKMLGSLSDHNSAVPEPTR-KITDPGAPKGY 300  
 QY 303 K---GEMEH-----AEVTVRSGQKAASLHELCAARVSEVLQNRVARETEVGVDFYAF 354  
 Db 291 KELYSPCL--DSFKGEWHE-----EVTVRSGQKAASLHELCA--AEVSEVLQNR 339  
 QY 304 KYSSGFLADBSKAGSLQAAGNFKCSATFALKKEGKENCYHSGSGSFTFDLQGS 363  
 Db 340 VERTTEVAVDFAVSYYDLAGVGLDAEKGSLLVGDPELIAKYVCTLETPQSPSCMDLTY-VSLLI 412  
 QY 364 FLAT-----ASFYTKAFPELEKGLSLIPAGRRYCG--EWSKLLIEYPTDEEYLR 416  
 Db 400 FSCMDLTY-VSLLIQFEGFPR-----SKVLLKTRKIDNVETSMALGA 440  
 QY 417 GYFSAAYTISLHBSLGIALDDESTITYASKGK--HLPDMALGA 461



## RESULT 9

S50463  
hypothetical protein YER005w - yeast (Saccharomyces cerevisiae)  
C:Species: Saccharomyces cerevisiae  
C:Date: 28-May-1993 #sequence\_revision 24-Feb-1995 #text\_change 19-Apr-2002  
C:Accession: S50463  
R:Dieterich, F.S.  
submitted to the EMBL Data Library, December 1994  
A:Description: The sequence of S. cerevisiae cosmids 9537, 9581, 9495, 9867, and lambda  
A:Reference number: S50433  
A:Accession: S50463  
A:Molecule type: DNA  
A:Residues: 1-630 <DIE>  
A:Cross-references: ENBL:U18778; NID:g603592; PID:g603597; GSPDB:GN00005; MIPS:YER005w  
C:Genetics:  
A:Gene: SGD:YND1; MIPS:YER005w  
A:Cross-references: SGD:S0000807  
A:Map position: 5R

Query Match 13.9%; Score 329.5; DB 2; Length 630;  
Best Local Similarity 26.9%; Pred. No. 7.8e-19;  
Matches 122; Conservative 72; Mismatches 163; Indels 97; Gaps 22;

```
QY 74 YGIMFDAGSTGRVHVQTRPRPTLTHE-----TPKAVKP 112
DB 10 FGIVIDAGSSGRIHVFKV---QDTESLLHATNQDSQSILOVSPHIHQEKDWTFK-LNP 64
QY 113 GLSAYADDDVEKSAQG---IRELLDVAKODIPDFWKATPLVKATAGLRLLPGEKAQK 167
DB 65 GLSSP---EKPDQAYKSHIKPLDFAKNIIPESHNSCFVFIQATAGMLLPQDIQSS 120
QY 168 LLOQKVEFK-ASPLVIGDDC---VSMNGTDEGVSAMITINFLGSLK--TPGGS---S 218
DB 121 ILDGLCQGLKHPAEFLV-EDCSAQIQVIDGTEGLYGLNLYGHENDYNPEVSDHFT 179
QY 219 VGMLDLGGSTQIAFLPRVEGLQASPPGYLTA-LRMFN---RTYKLYSYSVLGLGMSA 274
DB 180 FGMMDGGASTQIAFAPHDSGEIARHDDIATILRSVNGDIQKWDVFSVTLGFGANQA 239
QY 275 R---LAILGGVGEQPAKD-----GKEIVSPCLSPSPKGEWEHAEVTVRVSG----OKA 320
DB 240 RRYLAQLNLTLPENTNDYENDDFSRNLNDFCMFRGSDTFEFDKDTIFHLAGSGNVEQC 299
QY 321 AASLHELCAHV-----SEVLQVRHTE-EVGHVFPYAFSYDYDLAAGVLIDAEKGSGL 375
DB 300 TRKSIYPLLLKNMPCDDPECLFNGVHAPRIDFANDKFIGTSEWYTANDV----FKLGGEY 355
QY 376 VVGDFEIAAKYVCTRLETO-----POS-SPFSCMDLTVSVLLLOE-FGFPRS 420
DB 356 NFDKFSKSLREFGNSWTOILANSKGVYNSIPENFLKDACFKGNVNLILHEGDFMFI 415
QY 421 KV-----LKLTRKIDNVETSWALGAIFY 444
DB 416 DVDAENVNDRPLFQSVKEVEERELSWTLGRILLY 449
```

## RESULT 10

T16696  
hypothetical protein R07E4.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T16696  
R:Miller, N.  
submitted to the EMBL Data Library, October 1995  
A:Description: The sequence of C. elegans cosmid R07E4.  
A:Reference number: Z18561  
A:Accession: T16696  
A:Status: preliminary; translated from GB/EMBL/DD5J  
A:Molecule type: DNA  
A:Residues: 1-557 <ML>  
A:Cross-references: EMBL:U39652; NID:g1049390; PID:g1049394; PIDN:AA80403.1; CESP:R07E4

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:R07E4.4

A:Introns: 39/1; 67/1; 145/2; 192/1; 244/3; 317/3; 451/3; 488/2

Query Match 13.0%; Score 307; DB 2; Length 557;  
Best Local Similarity 23.5%; Pred. No. 4.6e-17;  
Matches 119; Conservative 88; Mismatches 178; Indels 122; Gaps 23;

```
QY 13 AKVAPLGLCVGVFIYVAVIKMHRATQAFASITRAAFGARWGQQAHSPLGTAADGHEV 72
DB 14 AMIFPP-----VIVFIYV-----EAHTSPKVIADQER 42
QY 73 FYGIMFDAGSTGRVHVQTRPRP-----RETPTL--THETFKAVPGLSAYADDDVEKSA 125
DB 43 SYGVICDAGSTGRVHVQTRPRP-----RETPTL--THETFKAVPGLSAYADDDVEKSA 102
QY 126 OGIRELLDVAKODIPDFWKATPLVKATAGLRLLP-----CEKAQKLLQKVEKFKAS 179
DB 103 EYLRPLMELAEHPIPEEKRYTFVFPFATAGMLLIDPEYVLQKAEVLEKLEKLPKIT 162
QY 180 PFLVGGDDCVSIMGTDGVSAMITINFL-----TGLSKTPGGS-----SVGMLDLGG 226
DB 163 SMQVLKHEHRIIEGKWEIYSIAVNVALGKFNKTATLDPGTPAHARQKTVGMIDMG 222
QY 227 GSTQIAP-LPRVETGLQASPPGYLTAIR---MFRNTYKLYSYSVLGL-----GLMSARLA 277
DB 223 ASAQIAPELDPDTPSPFSINVENINLGCDDSDLFK--YKLFVTTFILGYGVNEIRKYE 280
QY 278 ILGVEGEQPAKDGKELVSPCLSPSPKGEWEHAEVTVRVSGQ---KAAASLHELCAARVSE 334
DB 281 LLSKLKQD---NGTVIQDDCMPLNL-----HKTVTLE-NGENFVRRTGNTWNTCSNEVK 331
QY 335 VL-----QNRVHRTVEVK-----HYDFVAFS-YYVDLAAGVCLIDAEKGSGL 375
DB 332 LLNPESSEYCKAEAAKCYFAGVAPAPSPILPSNIEMFYSEYWYTHDVLGL-----GGQY 386
QY 376 VVGDFEIAAKYV---RTLETOFQSSPF-----SCMDLTVSVLLLOEFGFPRS 421
DB 387 DAENIAKKTQYCSKRWSTQASKKLYPRADBEERLTCQFKSAMITSVLHD-GFSVDK 445
QY 422 VLKLRKIDNV---EFTSWALGAIFY 445
DB 446 THNKFQSVSTIAGQEVQWALGAMIYHM 472
```

## RESULT 11

E86276  
hypothetical protein Fl1417.1 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001  
C:Accession: E86276  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
A:Authors: Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: E86276  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-405 <STO>  
A:Cross-references: GB:AE005172; NID:g7626666; PIDN:AAF43924.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 12.7%; Score 301; DB 2; Length 405;  
Best Local Similarity 25.4%; Pred. No. 9.1e-17;

Matches	104;	Conservative	65;	Mismatches	181;	Indels	60;	Gaps	13;
Qy	74	YGINFDAGSTGTRVHVQFTRPPREPTTLTHETFKAVKGLSAYADVDKSAQIGRELLD	133						
Db	4	FGYWFESGKP-----VDFDGE-----EHYASLKLSPGLSSYADNPEGASVSVTKLVE	50						
Qy	134	VAQODIPFFFWKATPLVLKATAGLRLLPGEKAOKLQKVEKFAKSPFLVGGDDCVIMMG	193						
Db	51	FAKGRIPXGKLKSSDRMLMATAGVRLDVPVQGIPLDVTTRVRLRSSGFKFQDEWATVIG	110						
Qy	194	TDSGVSAWITINFLTGSLTKPGSSVGMLLDGGSTQIAFLPRVEGTLOASPPGYLTALR	253						
Db	111	TDSGIYAWVANHAGSLGSDPLKTTGIVELGASAQVTFVPS-----EHVPFEFSRTIS	165						
Qy	254	MFNRTYKLYSYVYLGLGLMSARLAILGGVEG-QPAKDGLV-SPCL-----	298						
Db	166	YGNVSVTIYSHSFLDFQDAAAEKLLSLEQNSVAASTGCGIVEDPCTPKGYIDYTHSQXD	225						
Qy	299	SPSPKGEWEHAETVYRVSGGKAASLHELCAARVSEVLQNRVHRTBEVHV-----	349						
Db	226	SSGFLEB-----ESKFKASLQVQAAGDPTKCRSATLAWLQEGKENC-AYKHCSIGSTFTFN	280						
Qy	350	---DFVAFSYYYDLAAGVGLIDAEGGSLVVGDFEIAAKVYCVRTLETQTPOSSP----	402						
Db	281	IQGSFLATENFHTSKFFGLGEKWLSEMLACKRFCEGEWSKLKPYPTKDKYLHRYC	340						
Qy	403	MDLTY-VSLLLOBEGP-----PRSKVLKLTKIDNVETSWALGAIFHWIS	447						
Db	341	FSSAYIISMLHDSIGVALDDBERIKYASKAGC-ENIPLDWALGAFIINTDT	369						

RESULT 12  
I56242  
lymphoid cell activation antigen - human  
C:Species: Homo sapiens (man)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 18-Nov-2002  
C:Accession: I56242  
R:Maliszewski, C.R.; Delespesse, G.J.; Schoenborn, M.A.; Armitage, R.J.; Fanslow, W.C.;  
J. Immunol. 153, 3574-3583, 1994  
A;Title: The CD39 lymphoid cell activation antigen. Molecular cloning and structural cha  
A;Reference number: I56242; MUID:95015846; PMID:7930580  
A;Accession: I56242  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: mRNA  
A;Residues: 1-510 <RES>  
A;Cross-references: GB:S73813; NID:G765255; PIDN:AAB32152.1; PID:G765256  
C;Superfamily: nucleoside triphosphatase chromatin-associated

Query Match	12.1%;	Score	285.5;	DB 2;	Length	510;			
Best Local Similarity	26.2%;	Pred. No.	2.3e-15;						
Matches	114;	Conservative	68;	Mismatches	158;	Indels	95;	Gaps	20;

  

QY	72	VFYGMFDAGSTGTRVHVHVFQTRRPRETPILTH--ETFFKAVKPGLSAYADDVEKSNQGR	129
DB	47	VXYGIVLDAGSSHTSLYTRYPAEKENDGTGVVHQVEECRVKPGFISKFPQVNEIGIYLT	106
QY	130	ELLDDVAKODIPEDFWKATPLVKATAGLRLLPGSKAOKLQKYKEVFKAS----	PFLVGD 185
DB	107	DCWEAREAVIPRSQHQETFPVLTGATGNRLLRME--SEELADRVLDVVERSLSNYPF--	DF 163
QY	186	DCVSMNGTDEGVSAWITINFLTQSL-----	KTPGSSVGMGLDGGGGSTQIAF 233
DB	164	QGARIITQGEAGYCIWITINVLGKFSOKTRWFSIVPYETNNQETFGALDLGGASTQVTF	223
QY	234	LPRVETGLQASPGVLTALRPNFTYKLYSVSYLGLGMSARIALIIGVVEGQPAKD----	289
DB	224	VPQNQ--TIES--FDNALQFRLYGKDQVNYTHSFCLCYGKDQALWQL-----	AKD IQVA 273
QY	290	-GKEIIVSPCLSPFKGE-----WEHAETV-----YRVSG-----	QKAAASLHEI- 327
DB	274	SNEILLDPCTHEGYKKVNVNSDLKYPTCTKRFENMLPFQQFETGGTGNVQQCHQSILELF	333
QY	328	-----CAAR--VSEVLQNRVHRTTEEVKHVDVFAFSYYVYDLAAGVGLIIDAEKGSIV	376

D6		334	NTSYCPYSCAFNGIFLPLQG-----DFGAFSAFYVMKFLNLITSEKVSQEKV	382
QY		377	VGDFEIAAKYVCRTLETQPQS-----SPFSCMDLYV-SILLQEPGFPRS--KVJK	424
D6		383	T---EMMKFCQAQPNEEIKTSYAGVKESYLSEY-CPSGTIVLSLLQLGVHFTADSWEHIIH	438
QY		425	LTRKIDNVETSWALG	439
D6		439	FIGKIQQSDAGWTLG	453

RESULT 13

140856  
probable nucleotide phosphatase - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T40856  
R:Ramsperger, U.; Pohl, T.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, October 1999  
A:Reference number: Z21952  
A:Accession: T40856  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-572 <RAM>  
A:Cross-references: EMBL:AL121783; PIDN:CA857847.1; GSPDB:GN00068; SPDB:SPCC11E10  
A:Experimental source: strain 972h; cosmid c11E10  
C:Genetics:  
A:Gene: SPDB:SPCC11E10.05c  
A:Map position: 3

[illegible]

RESIST 14

hypothetical protein T18B16.150 - Arabidopsis thaliana  
T04439  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 30-Apr-1999 #sequence revision 30-Apr-1999 #text change 30-Apr-1999

C/Accession: T04439  
 R/By: Van, M.; Bene, V.; Rechmann, S.; Borikova, D.; Ansoorge, W.; Bancroft, I.; Mewes, H.W.  
 Submitted to the Protein Sequence Database, April 1998  
 A/Reference number: 215359

A/Accession: T04439  
 A/Molecule type: DNA  
 A/Residues: 1-1052 <REV>  
 A/Cross-references: EMBL:AL021687  
 A/Experimental source: cultivar Columbia; BAC clone T18B16  
 A/Genetics:  
 A/Map position: 4  
 A/Insertions: 79/3; 118/3; 217/1; 295/3; 396/3; 531/2; 815/3  
 A/Note: T18B16.150

Query Match 10.4%; Score 245.5; DB 2; Length 1052;  
 Best Local Similarity 25.3%; Pred. No. 1.2e-11;  
 Matches 116; Conservative 60; Mismatches 185; Indels 97; Gaps 20;

QY 55 WGGQAH-----PCTADGHE-----VYGMFDAGSTGTVHPQFT 93  
 DB 483 WGGGSSPSKESKGVPMGTPTSTRKLIKRAMTWMLFLPAFLVYISITGRAYVYQAS 542  
 QY 94 ---RPPETPTLTHETPKAV-----KPLSAVADDERKSAQIGIRELLDVAK 136  
 DB 543 INYKKSLSPLVMSLREGISRSRGRAYDMETPEPFDLVNNRGLKTAIKELIOMAE 602  
 QY 137 QDIPFDWKATPLVLTATAGIRLLPGEKAOGLQKVEKFAFPFLVGDVCSIMNGTDE 196  
 DB 603 KQIPKNAHRTSLVYATAGVRLRPAADSSWILGNVSIILAKSPFCRRRHWVXIIGSTER 662  
 QY 197 GVSAMITINFLTGLSL-KTPGGSSVGMDLGGGSGTOIAFLPRVEGTLOASPRVLTALRMF 255  
 DB 663 AYPFWTALNYOTSMGLPKPKATFGALDLGGSSIQVTF---EKERTHNETNLN-IRIG 717  
 QY 256 NRTYKLYSYSLGLGMSA-----RLAILGVYEGGPANDGK-ELVSPCLSPSEKEME 307  
 DB 718 SVNHLTAVSYLAGLMDAFDRSVVHLKLEPNVNSKDLIEGKLEMHGFCINSGYNGQYL 777  
 QY 308 HAEVTYVSGOKAASLHEL-----CAARVSE---VLONRVTRTEVGHVDFYAF 354  
 DB 778 CSQCSAVOGSKKSGVSIKLVGAPNMGESALAKAPCALPGYPR---PHGGFYAV 833  
 QY 355 S-----YYVDIAGVGLIDA-EKGGSLVGDDEFIAAKYVCRTELETPQSSPF---SCMD 404  
 DB 834 SGFVYVRRFFVLSABASLDVLEKG-----REFCDKAMQVART-SVSPQ--PFIHQYCFR 885  
 QY 405 LTYVSLIQEFGRSKVLTUKTK---LDNVTSMALG 439  
 DB 886 APYTVSLIRE-----GLYITDKQIITIGSGSITWTLG 916

## RESULT 15

C86276

7A19.33 protein - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 18-Nov-2002

C/Accession: C86276

R/Theologian, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 anson, N.F.; Hughes, B.; Hutzar, L.  
 Nature 408, 816-820, 2000

A/Authors: Hunter, U.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shin, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Ventier, J.C.; Davis, R.W.  
 A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A/Reference number: A86141; MUID:21016719; PMID:11130712

A/Accession: C86276  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-508 <STO>  
 A/Cross-references: GB:AE005172; NID:G5080800; PIDN:AMD39310.1; GSPDB:GN00141

C/Genetics:  
 A/Map position: 1  
 C/Superfamily: nucleoside triphosphatase chromatin-associated

Query Match 9.3%; Score 219; DB 2; Length 508;  
 Best Local Similarity 22.3%; Pred. No. 6.6e-10;  
 Matches 109; Conservative 83; Mismatches 196; Indels 100; Gaps 21;

QY 9 SLRAKVAVPLGLCVGFVYVAYIKWHATATQAFSTIRAPAGARWGGQAHSPGLGTAD 68  
 DB 26 SKRTSIIIFVIVACTTALGLFLG-----YSLRSGRRR----- 61  
 QY 69 GHEVFYGMFDAGSTGTVHVPQFTTPPETPTLT-HETFAVAV--DGLSAVADDERKS 124  
 DB 62 -VSLHVSVIDGSSGSTRVHVG-Y-RISGKVPFDGGEENYASLKLSPGLSAVADNPEGV 119  
 QY 125 AGGIRELLDVAKODIPFPWKATPLVLTATAGIRLLPGEKAOGLQKVEKFAFPFLV 184  
 DB 120 SESVTELVFPKGRVHKGLKKSDFRLMATGMRLLELPVQDILVTRVLRSSGDFDR 179  
 QY 185 DDCVIMNGTDEGVAMITINFLTGLSLKTPGGSSVGMDLGGGSGTOIAFLPRVEGTLOAS 244  
 DB 180 DEMASVISILENPDQIMKV-YMLGGL-----LIMSVRLEVFV-----STELV 223  
 QY 245 PRGYLTALMFRYTKLYSYSLGLGMSARLAILGVYEGGPANDGK-ELV-SPCLS--- 299  
 DB 224 PSFERTLAYGNVSYNLVSHSFLDPGODAAOEKLESLSYNSAANSTGIVPDPICPKGY 283  
 QY 300 -----PSFKGEMHAEVTVRVSG-----OKAASLHE--LCAARVSEV----- 335  
 DB 284 ILETNLOKDLPEFLADKGFRTLTQAAANPSECRBAALAMQEBKGTYYKRCSTGSIPT 343  
 QY 336 --LONRVHRTBEVKAVDPAVSYYVDIAGVGLIDAERKGGSLVGD--FEIAAKYVCRTE 391  
 DB 344 PMLQGSFLATE-----NFFHNSKFFGLGEKEWLSMILAGKRFCEHWSKLVKYP--- 394  
 QY 392 ETQPOSSPSCMDLNY-VSLIQERGF-RSKVLTUKTKI--DNVETSMALGAT----- 441  
 DB 395 TKFENLRYCPSASAYITISMLHDSGLVALDERIKYASKAGBBDIPLDWAALGFLNTAT 454  
 QY 442 --FHYIDS 447  
 DB 455 ATFDYSDN 462

Search completed: June 8, 2004, 09:50:35  
 Job time : 23 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 09:48:14 ; Search time 18 Seconds  
(without alignments)  
1319.111 Million cell updates/sec

Title: US-09-905-589A-2

Perfect score: 2364  
Sequence: 1 MKRISNHSRPAKVAIVPLG.....ALGAIHFYIDSLNRQSPAS 456

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues  
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2364	100.0	484	1	ENP6_HUMAN
2	2063.5	87.3	455	1	ENP6_RAT
3	1007	42.6	427	1	ENP5_MOUSE
4	999	42.3	428	1	ENP5_HUMAN
5	959	40.6	469	1	ENP5_MESAR
6	498	21.1	518	1	GDAL1_YEAST
7	485	20.5	454	1	APV_GOLTR
8	442.5	18.7	455	1	NTPA_PEA
9	343	14.5	485	1	YTA4_CABEL
10	329.5	13.9	630	1	YND1_YEAST
11	321.5	13.6	616	1	ENP4_HUMAN
12	320.5	13.6	613	1	ENP4_MOUSE
13	317.5	13.4	552	1	YBU4_CABEL
14	315.5	13.3	493	1	ENP1_CHICK
15	314.5	13.3	529	1	ENP1_HUMAN
16	299.5	12.7	510	1	ENP1_MOUSE
17	294	12.4	495	1	ENP2_MOUSE
18	289	12.2	494	1	ENP2_CHICK
19	287	12.1	495	1	ENP2_RAT
20	286	12.1	511	1	ENP1_RAT
21	285.5	12.1	510	1	ENP1_HUMAN
22	265	11.2	495	1	ENP2_HUMAN
23	265	11.2	513	1	ENP1_BOVIN
24	253	10.7	510	1	ENP1_PIG
25	140.5	5.9	628	1	NTP2_TOXGO
26	139	5.9	628	1	NTP1_TOXGO
27	128.5	5.4	592	1	NTP4_TOXGO
28	109.5	4.6	505	1	MOO_BSEFL
29	103.5	4.4	1516	1	CAIH_HUMAN
30	102.5	4.3	440	1	FTSA_BACSU
31	102	4.3	475	1	GATA1_THESA
32	102	4.3	623	1	DNAX_CAMGE
33	99.5	4.2	345	1	RLUB_PASMU

RESULT 1	ENP6_HUMAN	STANDARD	PRT	484 AA.
AC	075354	Q9UDD1		
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Botomucleoside triphosphate dihydrophosphate 6 (EC 3.6.1.6)			
DE	(NTPase6) (CD39 antigen-like 2).			
GN	ENTPD6 OR CD39L2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Keratinocytes;			
RX	MEDLINE=98341119; PubMed=9676430;			
RA	Chadwick B.P., Fritsch A.-M.;			
RT	"The CD39-like gene family: identification of three new human members			
RT	(CD39L2, CD39L3, and CD39L4), their murine homologues, and a member of			
RT	the gene family from Drosophila melanogaster."			
RU	Genomics 50:357-367(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21638749; PubMed=11780052;			
RA	Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,			
RA	Jones M., Stavridis G., Almeida J.P., Babbage A.K., Baggaley C.L.,			
RA	Bailey J., Barlow K.P., Bates K.N., Beard L.M., Beare D.M.,			
RA	Beasley O.P., Bird C.P., Blakey S.B., Bridgman A.M., Brown A.J.,			
RA	Buck D., Burrill W.D., Butler A.P., Carder C., Carter S.Y., Clee C.M.,			
RA	Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,			
RA	Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,			
RA	Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,			
RA	Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,			
RA	Graham D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E.,			
RA	Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.V., Howden P.J.,			
RA	Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,			
RA	Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,			
RA	Lehvaselaho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,			
RA	Martin V.L., Martin S.L., McConachie L.J., McKay K., McMurtry A.A.,			
RA	Miles S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,			
RA	Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,			
RA	Phillimore B.O.C.T., Prichard S.R., Plumb R.W., Ramsay H.,			
RA	Rice C.M., Ross M.T., Scott C.E., Seta H.K., Showken R., Sims S.,			
RA	Skuse C.D., Smith M.L., Soderlund C., Steward C.A., Suleston J.E.,			
RA	Swann R.M., Symamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,			
RA	Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,			
RA	Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,			
RA	Willing L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,			
RA	Rogers U.;			
RT	"The DNA sequence and comparative analysis of human chromosome 20."			
RT	Nature 414:865-871(2001).			
CC	-!- FUNCTION: Might support glycosylation reactions in the Golgi			
CC	apparatus and, when released from cells, might catalyze the			
CC	hydrolysis of extracellular nucleotides. Hydrolyses preferentially			

34	99.5	4.2	470	1	DLN1_BACST	P11959 bacillus st
35	99.5	4.2	2411	1	DAB_DROME	P96081 drosophila
36	98.5	4.2	633	1	DNAX_AGRIS	P50019 agrobacteri
37	98	4.1	415	1	SAHH_SULTO	Q97510 sulfolobus
38	97.5	4.1	415	1	SAHH_METTA	Q58783 methanococc
39	97.5	4.1	548	1	LIP2_CANRU	P33946 candida rug
40	96	4.1	410	1	SAHH_THERO	Q97924 thermoplas
41	96	4.1	711	1	TOP1_MYCPN	P76032 mycoplasma
42	95.5	4.0	4074	1	PKHD_HUMAN	Q8C29 homo sapien
43	95	4.0	527	1	HSF8_LYCES	Q40152 lycopersic
44	95	4.0	549	1	LIP1_CANRU	P20261 candida rug
45	94.5	4.0	641	1	DNAX_XANNC	Q9pm0 xanthomonas



QY 121 VESKAGIRELLDYAKODIPEDFWKATPLVYKATAGRLLPGEKAOQLCKVZEVFPASP 180  
 DB 121 VESKAGIRELLDYAKODIPEDFWKATPLVYKATAGRLLPGEKAOQLCKVZEVFPASP 180  
 QY 181 FLVGDCCVSTNGGDEVSAMITINFLTSGSKTPGSGSVMLDGGSGTQIAFLPREVET 240  
 DB 181 FLVGDCCVSTNGGDEVSAMITINFLTSGSKTPGSGSVMLDGGSGTQIAFLPREVET 240  
 QY 241 LQASPPGYLTALRMFNRTYKLSYSYLGGLGMSARLAILGVSQAPAKDKEIVSPCLSP 300  
 DB 241 LQASPPGYLTALRMFNRTYKLSYSYLGGLGMSARLAILGVSQAPAKDKEIVSPCLSP 300  
 QY 301 SFKEMHEAETTVVSGQKAAASHLCAARVSVLQNRVHRTVEVHVPFYAPSTYYDL 360  
 DB 301 RFRKMEHAEVTVYRISQK-AVGYELCASRSEVLNKKVHREAAQHVFPFASYYYDL 359  
 QY 361 AAGVGLDAEKSGSLVWGDPEIANKVYCRLETPQSSPSCMDLTYSLLDGFPPRS 420  
 DB 360 AAGVGLDAEKSGSLVWGDPEIANKVYCRLETPQSSPSCMDLTYSLLDGFPPRS 419  
 QY 421 KVLKLRKIDNVTSMALGALFHYIDSILNRQKSPA 455  
 DB 420 KVLKLRKIDNVTSMALGALFHYIDSILNRQKSPA 454

RESULT 3  
 ENFS\_MOUSE STANDARD; PRT; 427 AA.  
 ID ENFS\_MOUSE 070214;  
 AC 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Ectonucleoside triphosphate diphosphonucleoside 5 precursor  
 DE (EC 3.6.1.6) (NTPdase) (Nucleoside diphosphatase) (CD39 antigen-like  
 4) (ER-UDPase).  
 GN ENTPD5 OR CD39L4.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=98119025; PubMed=9457681;  
 RA Chadwick B.P., Williamson J., Sheer D., Fritsch A.-M.,  
 RT "cDNA cloning and chromosomal mapping of a mouse gene with homology to  
 RT NTPase.";  
 RL Mamm. Genome 9:162-164(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, AND CHARACTERIZATION.  
 RC TISSUE=Liver;  
 RX MEDLINE=99298181; PubMed=10369669;  
 RA Trombetta E.S., Helentius A.,  
 RT "Glycoprotein regulation and nucleotide sugar utilization in the  
 RT secretory pathway: identification of a nucleoside diphosphatase in the  
 RT endoplasmic reticulum.";  
 RL EMBO J. 18:3282-3292(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Kidney;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
 RA Aikawa K., Iawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,  
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Flischnmann W., Gaasterland T., Gissi C., King B., Kochwa H.,  
 RA Kuehl P., Lewis S., Masuno Y., Nikaido T., Pesole G., Quackenbush J.,  
 RA Schrimm L.M., Staudt F., Suzuki R., Tomita M., Wagner D., Maehiro T.,  
 RA Sakai K., Okido T., Furuno M., Kono H., Baldarelli R., Barsh G.,  
 RA Blake J., Botfield D., Bonunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli U., Mombereis P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasai H., Sato K., Schenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.,  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Kidney;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Tashiro S., Carninci P., Prange C.J.,  
 RA Rana S.S., Loggiano N.A., Peters G.J., Abramson R.D., Mullen S.J.,  
 RA Bosak S.A., McKernan K.J., Malek J.A., Gharatne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Munz D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schermer A., Schein J.E., Jones S.J.W., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- FUNCTION: Likely to promote glycosylation reactions involved in  
 CC glycoproteins folding and quality control in the endoplasmic  
 CC reticulum. Hydrolyzes UDP, GDP AND IDP but not any other  
 CC nucleoside di-, mono- or triphosphates, nor thiamine  
 CC pyrophosphate.  
 CC -1- CATALYTIC ACTIVITY: A nucleoside diphosphate + H(2)O = a  
 CC nucleotide + phosphate.  
 CC -1- COFACTOR: Requires calcium and magnesium.  
 CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.  
 CC -1- TISSUE SPECIFICITY: Ubiquitous.  
 CC -1- PTM: Glycosylated with high mannose N-linked glycans.  
 CC -1- MISCELLANEOUS: Optimal pH is neutral.  
 CC -1- SIMILARITY: Belongs to the GDA1 / CD39 NTPase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF006482; AAC05181.1; -;  
 DR EMBL: AJ238636; CAB4553.1; -;  
 DR EMBL: AK002618; BAB22234.1; -;  
 DR EMBL: BC015247; AAI15247.1; -;  
 DR MGD: MGI:1321385; Entrepd5.  
 DR InterPro: IPR000407; GDA1 CD39\_NTPase.  
 DR Pfam: PF01150; GDA1 CD39\_1; NTPase.  
 DR PROSITE: PS01238; GDA1 CD39\_NTPase; FALSE\_NEG.  
 KW Hydroxylase; Transmembrane; Glycoprotein; Calcium; Magnesium;  
 KW Endoplasmic reticulum; Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 19 427  
 FT CARBOHYD 41 41  
 FT CARBOHYD 231 231  
 FT CONFLICT 350 350  
 FT CONFLICT 394 427  
 FT PTLAKSEHQRHDMWLGHLGSHLSPAVFSGHQLRBSSTSEAC

```
FT SEQUENCE 427 AA; 47101 MW; 156PVSQGVSTSTFSLGKAMPETR (IN REF. 1).
SQ Best Local Similarity 50.0%; Score 1007; DB 1; Length 427;
Matches 213; Conservative 62; Mismatches 133; Indels 18; Gaps 6;

QY 37 ATATQAFITRAAPGA-----RWGQ-----QAHSPGLCTAAGDHEVYFIMPDAGSTG 84
DQ 2 ATSWGAFFVLLIACVGSSTVFYREQQTFVFGVFLSSMCPINVSAG---TFYIMPDAGSTG 58
QY 85 TRVHPQFT-RPPREPTTLTHTFKAQKPLGSAVADDEKSAQGIKRELLDVAQDIPDF 143
DQ 59 TRHVTVFQKTAGQLPFLEGEIFDSVKEFLSAFVDQPKQGAETVQELLEVAKDSIPRSH 118
QY 144 WKATPVLVKTAGLRLLPQKQKLLQKVEFKASPLVGDGVSIMNGTDEGVSAWIT 203
DQ 119 WERTPVVLTATAGLRLLPQKQKALLLEVEEFKNSPLVPGDVSIMDGSVEGFLAWVT 178
QY 204 INFLTSLKTPGSSVGMGLDGGSTQIAFLPRVGTIQAASPPGYLTALRMFNRITYKLS 263
DQ 179 VNFLLQQLHGRGQETVGTLLDGGASTQITFLQFQKTEQTPRGVLTSEFNFSTFKLYT 238
QY 264 YSYLGGLMSARLAILGGVGPQAKGKELVSPCLSPSKGWEHAETVYRVSGQKAAAS 323
DQ 239 HSYLGFLKAARLATLGALEAK-GTDGHTFRSACLPRMLEASWIFGVKYQYGGNQGEM 297
QY 324 LHELCAARVSEVLONRVHTESKVDVFAFYVYVDLAAGVGLIDAEXGSLVVDGFIA 383
DQ 298 GFEPCEAEVLRVVGKQHPQEEVGRSAFAFYFYVDRAADTHLDYKGGVLUKVEDFRK 357
QY 384 AKVYCTLETQPOSSPFSCMDLTYVSLLLQE-FGFPKRSVLMKTRKIDNVETSWALGAIF 442
DQ 358 AREVCNLSGSSFGSSFFLQMDLTYITALLKDGFGADGTLQLTKKVNNIETGVALGATF 417
QY 443 HYIDSL 448
DQ 418 HLLQSL 423

RESULT 4
ENPS_HUMAN
ID_ENPS_HUMAN STANDARD; PRT; 428 AA.
AC 075356;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ectonucleoside triphosphate diphosphohydrolase 5 precursor
DE (EC 3.6.1.6) (NTPases) (Nucleoside diphosphatase) (CD39 antigen-like
DE 4) (ER-UDPase).
GN ENTPD5 OR CD39L4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leukemia;
RX MEDLINE=98341119; PubMed=9676430;
RA Chadwick B.P., Frisch A.-M.;
RT "The CD39-like gene family: identification of three new human members
RT (CD39L2, CD39L3, and CD39L4), their murine homologues, and a member of
RT the gene family from Drosophila melanogaster.";
RL Genomics 50:357-367(1998).
CC -!- FUNCTION: Likely to promote reglycosylation reactions involved in
CC glycoproteins folding and quality control in the endoplasmic
CC reticulum. Hydrolyzes UDP, GDP AND IDP but not any other
CC nucleoside di-, mono- or triphosphates, nor thiamine pyrophosphate
CC (By similarity).
CC -!- CATALYTIC ACTIVITY: A nucleoside diphosphate + H(2)O = a
CC nucleotide + phosphate.
CC -!- COFACTOR: Requires calcium and magnesium.
CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen (By similarity).
```

```
CC -!- TISSUE SPECIFICITY: Expressed in adult liver, kidney, prostate,
CC testis and colon. Much weaker expression in other tissues.
CC -!- SIMILARITY: Belongs to the GDAI / CD39 NTPase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC ENBL; AF039918; AAC39885.1; -
CC GenBank; HGNC:3367; ENTPD5.
CC MM; 603162;
CC InterPro; IPR000407; GDAI_CD39_NTPase.
CC Pfam; PF01150; GDAI_CD39; 1.
CC PROSITE; PS01238; GDAI_CD39_NTPASE; FALSE NEG.
CC Hydrolase; Transmembrane; Glycoprotein; Calcium; Magnesium;
CC Endoplasmic reticulum; Signal.
CC SIGNAL 1 20 POTENTIAL
CC CHAIN 21 428 ECTONUCLEOSIDE TRIPHOSPHATE
CC DIPOHOSPHOHYDROLASE 5.
CC CARBOHYD 42 42 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 232 232 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 428 AA; 47517 MW; 830437A155DE4DDD CRC64;
SQ
Query Match 42.3%; Score 999; DB 1; Length 428;
Best Local Similarity 52.4%; Pred. No. 8.8e-72;
Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;

QY 62 PLGTAADGHEVFVIGIMFDAGSTGTRVHVFOFT-RPPREPTTLTHTFKAQKPLGSAVAD 120
DQ 40 PINVSA---STLYGIMFDAGSTGTRIHVTVFQKMPGQLPILGEVEFDSVKPGLSAFVDQ 96
QY 121 VEKSAQGIKRELLDVAQDIPDFWKATPLVLKATAGLRLLPQKQKLLQKVEFKAS 180
DQ 97 PKQAEVTVQGLLEVAKDSIPRSHWKTTPVVLKATAGLRLLPQKQKALLFEVKEIFRSP 156
QY 181 FLVGDDCVSIMGTDGVSAWITINFLTGLKTPGSSVGMGLDGGSTQIAFLPRVEGT 240
DQ 157 FLVPKGSVSIIMDGSDEGLAWTVNFLTQQLHGRQETVGTLLDGGASTQITFLQFQKT 216
QY 241 LQASPPGYLTALRMFNRITYKLSYVGLGLMSARLAILGGVGPQAKGKELVSPCLSP 300
DQ 217 LEQTPRGYLTSEFNFSTYKLYTHSYLGLFGLKAARLATLGALETE-GTDGHTFRSACLPR 275
QY 301 SFKGEWEHAETVYRVSGQKAAASLHELCAARVSEVLQNRVHRTVEVKHVDVFAFYVDL 360
DQ 276 WLEAEWTFGGVYQYGGNQGEGVGFPCYAEVLRVVRGKLHQPEEVQSGSYAFSYIDR 335
QY 361 AAGVGLIDAEXGSLVVGDFEIAAKYVCTLETQPOSSPFSCMDLTYVSLLLQE-FGFP 419
DQ 336 AVDTMDIDYEKGGILKVEDEFERKAREVCNLENFTSGSPFLQMDLSYITALLKDGFGAD 395
QY 420 SKVLKTRKIDNVETSWALGAIFHYIDSL 448
DQ 396 STVQLTKKVNNIETGVALGATFHLQSL 424

RESULT 5
ID_ENPS_MESAU STANDARD; PRT; 469 AA.
AC 090YCB; 090YCB;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ectonucleoside triphosphate diphosphohydrolase 5 precursor
DE (EC 3.6.1.6) (NTPases) (Nucleoside diphosphatase) (CD39 antigen-like
DE 4) (ER-UDPase) (Proto-oncogene cph).
GN ENTPD5 OR CD39L4 OR CPH.
OS Mesocricetus auratus (Golden hamster).
```





```
FT TRANSMEM 10 24 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT DOMAIN 25 518 LUMENAL (POTENTIAL).
FT CARBOHYD 41 41 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 335 335 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 518 AA; 56821 MW; 9A61182D5ED22ADF CRC64;

Query Match 21.1%; Score 498; DB 1; Length 518;
Best Local Similarity 33.1%; Pred. No. 7.1e-32;
Matches 144; Conservative 60; Mismatches 153; Indels 78; Gaps 15;

QY 65 TADGHEVYGMDFDAGSTGTRVHVQF---TRPRETETLTHENFKAVKPGLSAYADV 121
DB 86 TCSEEHK--YVIMIDAGSTGSRVHIYKFDVCHSP---PTLDEKFDMLPEGLSSFDYDS 139
QY 122 ESKAQGIRELLDVAKODIPDFFWKATPLVLKATAGLRLLPGEKAQKLQKVE--VFKASP 180
DB 140 VGAANSLDPLLVAMNYVPIKARSTPVAVKATAGLRLLGDAKSKILSAVRDHLEKDYV 199
QY 181 F-LVGDDCVSINMGTDGVSAMTINFLGSLKTPGGS--SVGMLDLGGSTQIAFLPRV 237
DB 200 FPVEGGGVSIIMGDEGVFAMITNLLGNIGANGPKLPTAAVFDLGGGSTQIVFEPTF 259
QY 238 EGTQASPPGYLTALMFMNRTKLYSYGLGLMSARLAILGGVEGOPAKDGK----- 291
DB 260 PINEKMDVGEHKFDLKFGENYTYQFSLGYGLKEGRKNVSVLVENALKDGKILKGDN 319
QY 292 ----ELVSPCLSPSFKGEW-----HAETVTV-----SQKAAASLHELCAARVSEV 335
DB 320 TKTHQLSSPCLPKVKNATNEKVLESKETVYDFDIFGDPFSGAQ-----CREFLTDEI 371
QY 336 LQ-----NRVHRTTEEKVV-----DFVAFSYVYDLAGVGLIDAEGKGLVWGD 379
DB 372 LNKDAQCCQPPCFNGVHGPSLVRTKESNDIYISFYFDRTRPLGM-----PLSFTLINE 426
QY 380 FETAAYKIVKRTLET-----QPQSPFFSCMDITY-VSLLQEFQPPRSKVVLK 425
DB 427 LNDLARIIVKGEETWNSVFSFGAGSLDELESDFCLDSFQVSLIHTGYDIFLORELRT 486
QY 426 TRKIDNVEITSWALGA 440
DB 487 GKIANKEIGWCLGA 501

RESULT 7
APY_SOLTU STANDARD; PRT; 454 AA.
AC P80536; Q43164;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Apyrase precursor (EC 3.6.1.5) (ATP-diphosphatase) (Adenosine
DE diphosphatase) (ADPase) (ATP-diphosphohydrolase).
GN R80P1.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC lamids; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 59-160; 236-253 AND 332-345.
RC TISSUE-Tuber;
RX MEDLINE=96158985; PubMed=8579614;
RA Handa M., Guidotti G.;
RT "Purification and cloning of a soluble ATP-diphosphohydrolase
RT (apyrase) from potato tubers (Solanum tuberosum).";
RL Biochem. Biophys. Res. Commun. 218:916-923(1996).
RN [2]
RP SEQUENCE OF 42-54; 68-95 AND 236-253.
RC STRAIN=cv. Desiree;
RX MEDLINE=96355615; PubMed=8703025;
RA Vasconcelos E.G., Ferreira S.T., de Carvalho T.M.U., de Souza W.,
```

```
RA Kettlun A.M., Mancilla M., Valenzuela M.A., Verjovski-Almeida S.;
RT "Partial purification and immunohistochemical localization of ATP
RT diphosphohydrolase from Schistosoma mansoni. Immunological cross-
RT reactivities with potato apyrase and Toxoplasma gondii nucleoside
RT triphosphate hydrolase.";
EL J. Biol. Chem. 271:22139-22145(1996).
CC -!- FUNCTION: Catalyzes the hydrolysis of phosphoanhydride bonds of
CC nucleoside tri- and di-phosphates.
CC -!- CATALYTIC ACTIVITY: ATP + 2 H2O = AMP + 2 phosphate.
CC -!- COFACTOR: Calcium.
CC -!- SUBCELLULAR LOCATION: Membrane-associated (Probable).
CC -!- PTM: The N-terminus is blocked.
CC -!- SIMILARITY: Belongs to the GDAL / CD39 NTPase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U58597; AAB02720.1; -.
DR PIR; JC4616; JC4616.
DR InterPro; IPR000407; GDAL_CD39_NTPase.
DR Pfam; PF01150; GDAL_CD39; 1.
DR PROSITE; PS01238; GDAL_CD39_NTPASE; 1.
KW Hydrolase; Transmembrane; Calcium; Signal.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 454 APYRASE.
FT TRANSMEM 426 446 POTENTIAL.
FT CARBOHYD 151 151 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 454 AA; 50041 MW; 9D9EFE431DA2F52F CRC64;

Query Match 20.5%; Score 485; DB 1; Length 454;
Best Local Similarity 31.9%; Pred. No. 6.4e-31;
Matches 138; Conservative 68; Mismatches 183; Indels 44; Gaps 12;

QY 59 AHSPLGTAAAGHE-VFYGMFDAGSTGTRVHVQFTRPPRETPTLTH-ETFKAVKPGLSA 116
DB 30 AQPLRRLHLLSHSEHYAVIFDAGSTGSRVHVFRFDEKLLPIGNIEYFMATEPGLSS 89
QY 117 YADVEKSAQGIREDLDVAKODIPDFFWKATPLVLKATAGLRLLPGEKAQKLQKVEF 176
DB 90 YADPKAANSLFPLLDGAEVVPQELQSETPLELGATAGLRMLKGDAEKILQAVRNLV 149
QY 177 K-ASPELVGDDCVSINMGTDGVSAMTINFLGSLKTPGGSVGMGLDGGSTQIAFLP 235
DB 150 KNGTFHKKQWVITLDGTQEGSYMMAINLLGNLQKYKSTATIDLGGSVQVAYAI 209
QY 236 RVETGLQA----SPPGYLTALRMFNRTKLYSYGLGLMSARLAILGGVEGOPAKDGK 291
DB 210 SNEQFAKAPQNEDEGPYVQCKHLSKDYNLVHSLYNYGOLAGRAEIF-----KASR 261
QY 292 ELVSPCLSPSFKGEWHAETVTVSGKAAASLHELCAARVSEVLO-NRVHRTTEEK--- 347
DB 262 NESNFCALGECGGYISYGGVDYKVPKPKGSS-WKCRRLTRHALKINKNCTEETFG 320
QY 348 -----HVDYAFSYVYDLAGVGLIDAEGKGLVWGD-DFEIAAKYVCR----- 390
DB 321 VNWGGGDDGQKNTHASSFFDYIGAQGVIVDTKFPSSALAKPIQVYLNAAKACQTNVADIKS 380
QY 391 --LETQSSPFFSCMDITY-VSLLQEFQPPRSKVVLKTRKIDN----VETSWALGAIFH 443
DB 381 IFPKTDQRNIPYLCMDLIYETLLVDGFGNLPKHEITVIHDVQKNYLVGAAPLGAID 440
QY 444 YIDSLNRQKSPAS 456
DB 441 LVSTTNKIRVAS 453

RESULT 8
```



```

Db 251 NDCSPRGNLRLIGETVNTGTGEW-----DVCLAQVSLGDKAQP 291
QY 336 -----LQNRVHRTTEEVKHYDFAFSYVYDLAAGVGLIDAEGKGSVLVGDFFEAAXV 387
Db 292 SCNPPTCFELRNVIAPSVNLSTVQLYGFSEYWTTSNFG-----SGGEYHOKFTDEVKRY 346
QY 388 CR-----TLTQOSPPF-----SCMDTYVSLLLQEPGPRSKVLKTR---KID 430
Db 347 CQKMDNDIQDGFKRNEFPNADIERLGTNCFKAAVTSVLHD-GFNVNDRKHLFQSVLKA 405
QY 431 NVETSWALGAIFHYIDSL 448
Db 406 GEEMQWALGNMLYHKSOL 423

RESULT 10
YND1_YEAST
ID YND1_YEAST STANDARD; PRT; 630 AA.
AC P40009;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Golgi apyrase (EC 3.6.1.5) (ATP-diphosphatase) (Adenosine
DE diphosphatase) (ADPase) (ATP-diphosphohydrolase) (Golgi nucleoside
DE diphosphatase).
GN YND1 OR YER005W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=S288C;
RX MEDLINE=99340091; PubMed=10409709;
RA Gao X.D., Kaigorodov V., Jigami Y.;
RT "YND1, a homologue of GDAL, encodes membrane-bound apyrase required
RT for Golgi N- and O-glycosylation in Saccharomyces cerevisiae.";
RL J. Biol. Chem. 274:21450-21456(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313264; PubMed=9159868;
RA Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,
RA Araujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,
RA Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G.,
RA Hunick-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,
RA Lin D., Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P.,
RA Oh C., Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,
RA Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome V.";
RL Nature 387:78-81(1997).
CC -I- FUNCTION: Catalyzes the hydrolysis of phosphoanhydride bonds of
CC nucleoside tri- and di-phosphates. Has equal high activity toward
CC ADP/ATP, GMP/GTP, and UTP/UTP and approximately 50% less toward
CC CDP/CTP and thianine pyrophosphate. Has no activity toward GMP.
CC Required for Golgi glycosylation and cell wall integrity.
CC -I- CATALYTIC ACTIVITY: ATP + 2 H(2)O = AMP + 2 phosphate.
CC -I- PATHWAY: Glycosylation.
CC -I- SUBCELLULAR LOCATION: Golgi; membrane-bound.
CC -I- SIMILARITY: Belongs to the GDAL / CD39 NTPase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; AP203695; AAF17573.1; -.
CC ENBL; U18778; AAB64538.1; -.
CC PIR; S50463; S50463.

```

```

DR Germonline; 139085; -.
DR SGD; S000807; YND1.
DR GO; GO:0017110; F: Nucleoside diphosphatase activity; IDA.
DR GO; GO:0006486; P: Protein amino acid glycosylation; IMP.
DR InterPro; IPR000407; GDAL_CD39_NTPase.
DR Pfam; PF01150; GDAL_CD39_1.
DR PROSITE; PS01238; GDAL_CD39_NTPase; 1.
KW Hydrolase; Transmembrane; Golgi stack; POTENTIAL.
FT DOMAIN 1 500 LUMENAL (POTENTIAL).
FT TRANSXEM 501 517 POTENTIAL.
FT DOMAIN 518 630 CYTOPLASMIC (POTENTIAL).
FT SEQUENCE 630 AA; 71851 MW; 02F8D24A78212544 CRC64;

Query Match 13.9%; Score 329.5; DB 1; Length 630;
Best Local Similarity 26.9%; Pred. No. 2.2e-18;
Matches 122; Conservative 72; Mismatches 163; Indels 97; Gaps 22;

QY 74 YGIMFDAGSTGTRVHVHFQFTRPPRTPTLTHE-----TFKAVKP 112
Db 10 FGIVIDAGSSGSIHVFKW---QDTESLLHATNQDSQILQSVPHIQEKDWTFF-LNP 64
QY 113 GLSAYADDVEKSAQG-----IRELLVAKODIPDFEWKATPLVLKATAGLLPGEKAQK 167
Db 65 GLSSF-----EKKPDQDAYKSHIKPLLDFAKNIPESHWSGCFVIQATAGMLLPQDIQSS 120
QY 168 LLOKVZEVFK-ASPFVLVGDCC---VSIMMGTDREVSATWITINFLTGLSK--TPGGS---S 218
Db 121 ILDLGCLQGLKHPAEFLV-EDCSAQIQVIDGETEGYGLNLYGHENDYNEVSDHFT 179
QY 219 VGMLDLGGSTQIAFLPRVEGTLOAQPFGVLTG-IRMFN---RTYKLYSYVYLGLGMSA 274
Db 180 FGFMDMGASTQIAFAPHDSGETIARHDDIATIFLSVNGDLQKWDVVFVSTLWGFQANQA 239
QY 275 R---LATLGGVEGQAKD-----GKELVSPCLSPSKGMEHAETVTVRVSG-----QKA 320
Db 240 RRYLAQLINTLPENTYNDYNDPFCNPRGSSDTFFKDTIPIHAGSGNYEQC 299
QY 321 AASLHELCAARV---SEVLQNRVHRT-EVKHVDYAFSYVYDLAAGVGLIDAEGKGS 375
Db 300 TKSITYPILLKNMPCDDDEPCLFNGVHAPRIDFANDKFIETSEYWTYANDV---FKLGGEY 355
QY 376 VVGDFEIAAKYVCRTLETO-----PQS-SPFSCMDLTVYVSLLLQ-EGFPRS 420
Db 356 NDFKFSKSLREFCNSNWTQLANSKGVNISPENFLKDACFKGNVWVLLIHGEGFDPRI 415
QY 421 KV-----LKLTRKIDNVETSWALGAIFHY 444
Db 416 DVDAENVNDRPLFQSVKEVERELSLTIGRILLY 449

RESULT 11
ENP4_HUMAN
ID ENP4_HUMAN STANDARD; PRT; 616 AA.
AC Q9Y27; O15092;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ectonucleoside triphosphate diphosphohydrolase 4 (EC 3.6.1.6)
DE (NTPDase4) (Uridine-diphosphatase) (UDPase) (Lysosomal apyrase-like
DE protein of 70 kDa).
GN LYSALI OR ENTPD4 OR LALP70 OR KIAA0392.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND CHARACTERIZATION.
RC TISSUE=Brain;
RX MEDLINE=98225229; PubMed=95566635;
RA Wang T.-P., Guidotti G.;
RT "Golgi localization and functional expression of human uridine
RT diphosphatase.";
RL J. Biol. Chem. 273:11392-11399(1998).

```

RE SEQUENCE FROM N.A. (ISOFORM 1).  
 RX TISSUE=Pancreatic adenocarcinoma;  
 RX MEDLINE=99329914; PubMed=10399803;  
 RA Biededick A., Rose S., Elsaesser H.-P.;  
 RT "A human intracellular apyrase-like protein, LALP70, localizes to  
 RT lysosomal/autophagic vacuoles.";  
 RL J. Cell Sci. 113:2473-2484(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND ALTERNATIVE SPLICING.  
 RX MEDLINE=20317099; PubMed=10856452;  
 RA Biededick A., Kossan C., Kunz J., Elsaesser H.-P.;  
 RT "First apyrase splice variants have different enzymatic properties.";  
 RT J. Biol. Chem. 275:19018-19024(2000).  
 RN [4]  
 RP SEQUENCE OF 59-616 FROM N.A. (ISOFORM 2).  
 RL DNA Res. 4:151-150(1997).  
 CC -1- FUNCTION: Hydrolyzes preferentially nucleoside 5'-diphosphates,  
 CC nucleoside 5'-triphosphates are hydrolyzed only to a minor extent.  
 CC The order of activity with different substrates is UDP > GDP =  
 CC CDP = TDP, AMP, ADP, ATP and UMP are not substrates. Preferred  
 CC substrates for isoform 2 are CTP, UDP, CDP, GTP AND GDP, while  
 CC isoform 1 utilizes UTP and TTP.  
 CC -1- CATALYTIC ACTIVITY: A nucleoside diphosphate + H(2)O = a  
 CC nucleotide + phosphate.  
 CC -1- COFACTOR: Isoform 1 requires calcium and to a lesser extent,  
 CC magnesium. Isoform 2 is equally dependent on calcium and  
 CC magnesium.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC Localized in the Golgi and autophagic vacuoles/Lysosomes.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1; Synonyms=LALP70;  
 CC IsoId=Q9Y227-1; Sequence=Displayed;  
 CC Name=2; Synonyms=LALP70V;  
 CC IsoId=Q9Y227-2; Sequence=VSP\_003614;  
 CC -1- TISSUE SPECIFICITY: Ubiquitous. Highest expression in testis and  
 CC lowest in bladder.  
 CC -1- SIMILARITY: Belongs to the GDA1 / CD39 NTPase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL; AF016032; AAC17217.1; -;  
 DR EMBL; AJ131358; CAB40445.1; -;  
 DR EMBL; AJ246168; CAB45685.1; JOINED.  
 DR EMBL; AJ246166; CAB45655.1; JOINED.  
 DR EMBL; AJ246167; CAB45655.1; JOINED.  
 DR EMBL; AJ246168; CAB45655.1; JOINED.  
 DR EMBL; AJ246169; CAB45655.1; JOINED.  
 DR EMBL; AJ246170; CAB45655.1; JOINED.  
 DR EMBL; AJ246171; CAB45655.1; JOINED.  
 DR EMBL; AJ246172; CAB45685.1; JOINED.  
 DR EMBL; AJ246173; CAB45655.1; JOINED.  
 DR EMBL; AJ246174; CAB45655.1; JOINED.  
 DR EMBL; AJ246175; CAB45655.1; JOINED.  
 DR EMBL; AJ246176; CAB45655.1; JOINED.  
 DR EMBL; AB002390; BAA21575.1; -;  
 DR GeneW; HGNC:14573; LYSAL1.  
 DR MIM; 607577; -;  
 DR GO; GO:0030173; C:integral to Golgi membrane; IDA.

Query Match	13.6%	Score 321.5;	DB 1;	Length 616;
Best Local Similarity	24.9%;	Pred. No. 9.3e-18;		
Matches 132;	Conservative	69;	Mismatches 207;	Indels 123; Gaps 20;
QY	9	SLRVAKXAVPLGCGVGVFIYAVY--KMRHATATQAFSS-TRAPAGARWQQAHSPLGT	65	
Db	30	NLRQIMNTISVLAANVSLTFSVVIIRNKKYGLTKDKF-----QRIARPTDIAT	80	
QY	66	AADGHEVFGYIMPDAGSTGRVHFQET--PREPTLTHERFK-----AVKGLS	115	
Db	81	DTNNPNVNYGIIVDCGSSGRVFCYCPRHNGNPHDLIDIKQMDKRRKPVYMKIPGIS	140	
QY	116	AYADDEKSAQGITELLDVAKODIPDFPKATPLVLAATAGRLTLPKKAQKLLQK-VKE	174	
Db	141	EFATSPKSVSDYISPLINFAAEHPRAKHKETPYILCTAGMRILPSQQALIEDLTD	200	
QY	175	VFKASPFVCGDDCSINMGTEDEGVASMTTINFLTG-----SLTKPGSS---	218	
Db	201	IPNMFDELPSHAEVLSGKQGVYANIGINFLVGREHIEDDEDAVEVNIPESEGEA	260	
QY	219	-----VGMLDGGSGTQIAF-LRVEGTLQASPPGYL-TALRNFN-----RTYKL	261	
Db	261	IVRRRTAGLIDMGVSTQIAEVEVKTVSFPASSQOEVAKNLLAEFNGCDVHQTEHYRV	320	
QY	262	YSYSYVGLGGLMSARL-----ATGGVEG-----QPAKDGKELVSPCLSPSEFX	303	
Db	321	VVAIFLFGGNAARQRYEDRIFPANTIQCNRLIGQOTGLTPMPYLD-----PCLPLDIK	374	
QY	304	GEWHAETVYRVSGQKAASLHEDCAARVSEVLONRVHRTVEV-----KRV	349	
Db	375	DEIQNGQTVLRG---TGDFDLC---RETIQPMKNKINETQSLNGVQPIPHQNS	426	
QY	350	DFYAFSTYYDLAAGVGLDAEKGSLVVGDEIAAKYCR-----LETO	394	
Db	427	EYFSESEYYCTEDV---LIMGDDYNAAKTKYAKQYCAKWKSLRERDFGLYASHAD	482	
QY	395	QSSPFSQMDLTYV-SLLDGEFGFP-SSKVLKLTIRKIDNVETSWALGAIFH	443	
Db	483	LHRKLYQCFKSNMFEVYHRRGFSFPVNYKSLKTLQYVDKXVQNTLGLILY	533	

RESULT 12

ENP4\_MOUSE STANDARD; PRT; 613 AA.

AC 09DBT4;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last Sequence Update)

DT 15-MAR-2004 (Rel. 43, Last Annotation Update)

DE Ectomycloestite triphosphate dihydrophosphatase 4 (EC 3.6.1.6)

DE (NTPDase4) (Uridine-diphosphatase) (UDPase) (Lysosomal apyrase-like protein of 70 kDa).

DE LYSAL1 OR ENTPD4 OR LALP70.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC STRAIN=C57BL/6J; TISSUE=Lung;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Frieschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Roffelli D., Bojunga N., Carrincci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordore P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Togo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
RA Hayashizaki Y.,  
RT "Functional annotation of a full-length mouse cDNA collection";  
RL Nature 409:685-690(2001).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RC TISSUE=Brain, and Mammary gland;  
RX MEDLINE=22398257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.R., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carrincci P., Prange C.,  
RA Rana S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McSwain P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinska M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [3]  
RP ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.  
RX MEDLINE=20317099; PubMed=10858452;  
RA Biederbeck A., Kusan C., Kunz J., Elsasesser H.-P.;  
RT "First apyrase splice variants have different enzymatic properties.";  
RL J. Biol. Chem. 275:19018-19024(2000).  
CC -!- FUNCTION: Hydrolyzes preferentially nucleoside 5'-diphosphates,  
CC nucleoside 5'-triphosphates are hydrolyzed only to a minor extent  
CC (by similarity).  
CC -!- CATALYTIC ACTIVITY: A nucleoside diphosphate + H(2)O = a  
CC nucleotide + phosphate.  
CC -!- COFACTOR: Requires calcium and magnesium.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC Localizes in the Golgi and autophagic vacuoles/lysosomes (By  
CC similarity).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1; Synonyms=LALP70;  
CC IsoId=Q9DBT4-1; Sequences=Displayed;  
CC Name=2; Synonyms=LALP70V;  
CC IsoId=Q9DBT4-2; Sequences=VSP 003615;  
CC -!- TISSUE SPECIFICITY: Ubiquitous.  
CC -!- SIMILARITY: Belongs to the GDAL / CD39 NTPase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@sib-sib.ch).  
CC -----  
DR EMBL; AK004761; BAB23542.1; -;  
DR EMBL; BC006924; AAH06924.1; -;  
DR EMBL; BC043134; AAH43134.1; -;  
DR MGD; MGI:1914714; Lysall.  
DR InterPro; IPR000407; GDAL\_CD39\_NTPase.  
DR Pfam; PF01150; GDAL\_CD39\_1;  
DR PROSITE; PS01238; GDAL\_CD39\_NTPASE; FALSE NEG.  
KW Hydrolase; transmembrane; Glycoprotein; Calcium; Magnesium;  
KW Alternative splicing; Golgi stack; lysosome.  
FT DOMAIN 1 33 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 34 54 POTENTIAL.  
FT DOMAIN 55 559 LUMENAL (POTENTIAL).  
FT TRANSMEM 560 580 POTENTIAL.  
FT DOMAIN 581 613 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 404 404 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 407 407 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT VARSELIC 287 294 Missing (in isoform 2).  
FT /FTId=VSP 003615.  
SQ SEQUENCE 613 AA; 69745 MW; DES28F512ABEF52F CRC64;  
Query Match 13.6%; Score 320.5; DB 1; Length 613;  
Best Local Similarity 25.8%; Pred. No. 1.1e-17;  
Matches 120; Conservative 63; Mismatches 182; Indels 103; Gaps 17;  
Qy 65 TAADGHEVYGFIMFDAGSTGTTRVHVVFQFTR---PPRETTLTHETPK-----AVKPGI 114  
Db 80 TDTNPFVNYGIIVVDCGSSGRIFVYCWPRHNGPHDLDLDIRQMRDKNRKPVVMKIKPGI 139  
Qy 115 SAYADDERSAQOIRELLDVAKQDIPDFWKATPLVVKATAGLRLPLGKAKLQK-VK 173  
Db 140 SEFATSPKVSVDYISPLLSGFAAEHVPRAXHKEFYLILCTAGMRVLPSQQRAILSDLLT 199  
Qy 174 EVFKASPFVLGGDCVIMNGTDEGSVAVITINPLTG-----SLKTPGGSS-- 218  
Db 200 DIPHYDFLPSDSHAEVISQKQGVAVICINFLGRPEHIEDDEDAVVEVNIPEGSESE 259  
Qy 219 -----VGMLDIGGSGSTQIAP-LPVEGTQASPPGYL-TALRMFN-----RTYK 260  
Db 260 AIVKRTAGVLDMGVSTQIAYEVPQTQVSVFSSQQSEVAKNLLAEFNLGCDVHQTEHVYR 319  
Qy 261 LYSVSYGLGLMSARL-----AIIIGGVEGQPAKDGKELVSPCLSPSPKGEW 306  
Db 320 VYVATFLGFGNARQRYEDRLFASTVQKNRLGKQTG--LTPDAPLLDFCLPLDIKDEI 377  
Qy 307 EHAETVTVSGQKAAASLHLCARVSVLQNRVHRTTEV-----KHVDFY 352  
Db 378 QQNGQTLYLQG---TGDFDLC---RETLQPPMKNKTNETQTSINGVYQPPHFNQSEFY 429  
Qy 353 AFSYVYDLAAGVLDAEKGGSLVWGDPEIAKYVCRT-----LETQPOS 397  
Db 430 GFSEFYCTEDV-----LRMGGDYNARFTQAKDCATKWSILRERFDRLGLYASHADLHR 485  
Qy 398 SPFSCMDLTVV-SLLLQEFQFPFRS-KVLKILTRKIDNVETSWALGAIHF 443  
Db 486 LKQCQKSAWMEFVFKHGSFPVYTKNLTALQVYDKEVQWTLGAILY 533  
RESULT 13  
YBU4 CAEEL  
ID YBU4 CAEEL STANDARD; PRT; 552 AA.  
AC Q21815;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)



```
FT CARBOHYD 361 361 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 445 445 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 16 16 C -> W (IN REF. 2).
FT CONFLICT 21 21 I -> G (IN REF. 1; AA SEQUENCE).
SQ SEQUENCE 493 AA; 54034 MW; F14FF4C3AA2F3603 CRC64;

Query Match
Best Local Similarity 13.3%; Score 315.5; DB 1; Length 493;
Matches 114; Conservative 67; Mismatches 180; Indels 69; Gaps 15;

QY 65 TRADGHEVF-----YGLNFDAGSTGTRVHVQFQTRPPREPTLTH--ETTKAVKPGLSA 116
DQ 26 SAVDVKDFPPTGTYGLVDFDAGSTHTALYVYQNPADKENGCTGIVSVQVSECTVNGSGISS 85
QY 117 YADDVEKSAQGRRELLDVAQDIPDFWFKATPLVLKATAGLRLLPGEKAQKLLQKVKEVF 176
DQ 86 YADDPAGAGASLKPCLDKAMAVIPVEQWQOTPYLGATAGWRLLEQNSTKAEQVFAEVS 145
QY 177 KA-SPLVGDCCVSINMGTDGVSAMITNFI-----TGSUKTPGGSSV-GMLDLG 225
DQ 146 KAIREFPVDFRGAQILTKGNEEGSGWITVYLLLETKIFSPAGKWEHPQNTVELGALDLG 205
QY 226 GGSTOIAFLPRVEGTLQASPPGYLTALRMENRTYKLYSYVGLGLMSARLAILGGVEGQ 285
DQ 206 GASTQITFQGV--TIEDKNTSVL--FLYGTNHSYLSHTSVLCYQIQASKRLMAALH-- 259
QY 286 PAKDG---KELVSPCLSPFSKGEWHAEV-----TYRVSGOKAAASL 324
DQ 260 --QDGSYVQNIHPCVPGKRIITABIYDIPSPCVPTPSMLSPAQILTVTGTGNPAA--- 314
QY 325 HELCAARVSEVLQ-----NVRHTEEVKHV---DFYAFSVYVYDLAAGVGLIDAEKGSGL 375
DQ 315 ---CPTAILKFLNLCGANRTCGDGVYQPPVPGVGFAPAGFYTFPLNLTGQOQSLSHV 371
QY 376 VVGDEFLAAKYCYRTLETQPOSS-----PFSQMDLTYVSLLLQERGFPRS--KVLKLRKI 429
DQ 372 NATVWDFCNKWSSELVEFPQKHEHLHYCVVUGLYLLVLDVGYKDFDEHTSNHFSQKA 431
QY 430 DNVETSWALG 439
DQ 432 GNADIGWTLG 441

RESULT 15
ENP3_HUMAN
ID_ENP3_HUMAN STANDARD; PRT; 529 AA.
AC O75355; O60495;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-NAR-2004 (Rel. 43, Last annotation update)
DE Ectonucleoside triphosphate diphosphohydrolase 3 (EC 3.6.1.5)
DE (NTPDase3) (Ecto-ATP diphosphohydrolase) (ATPDase) (Ecto-apyrase)
DE (CD39 antigen-like 3) (HB6).
GN ENTPD3 OR CD39L3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retainocytes;
RX MEDLINE=98341119; PubMed=9676430;
RA Chadwick B.P., Frisch A.-M.;
RT "The CD39-like gene family: identification of three new human members
RT (CD39L2, CD39L3, and CD39L4), their murine homologues, and a member of
RL the gene family from Drosophila melanogaster.";
RL Genomics 50:357-367(1998).
RN [2]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Brain;
RX MEDLINE=98342144; PubMed=9675246;
```

```
RA Smith T.M., Kirley T.L.;
RT "Cloning, sequencing, and expression of a human brain ecto-apyrase
RT related to both the ecto-ATPases and CD39 ecto-apyrases.";
RL Biochim. Biophys. Acta 1386:65-78(1998).
RN [3]
RP REVISIONS.
RA Smith T.M., Kirley T.L.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP MUTAGENESIS OF TRP-187; ASP-219 AND TRP-459.
RX MEDLINE=99249795; PubMed=10231536;
RA Smith T.M., Lewis Carl S.A., Kirley T.L.;
RT "Mutagenesis of two conserved tryptophan residues of the E-type
RT ATPases: inactivation and conversion of an ecto-apyrase to an
RT ecto-NTPase.";
RL Biochemistry 38:5849-5857(1999).
RN [5]
RP MUTAGENESIS OF ARG-67; ARG-143; ARG-146; GLU-182; ASN-191; SER-224 AND
RL GLN-226.
RX MEDLINE=21197753; PubMed=11300774;
RA Yang F., Hicks-Berger C.A., Smith T.M., Kirley T.L.;
RT "Site-directed mutagenesis of human nucleoside triphosphate
RT diphosphohydrolase 3: the importance of residues in the apyrase
RT conserved regions.";
RL Biochemistry 40:3943-3950(2001).
CC -1- FUNCTION: Has a threefold preference for the hydrolysis of ATP
CC over ADP.
CC -1- CATALYTIC ACTIVITY: ATP + 2 H(2)O = AMP + 2 phosphate.
CC -1- COFACTOR: Requires calcium and magnesium.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- TISSUE SPECIFICITY: Expressed in adult brain, pancreas, spleen and
CC prostate. Moderate or low expression is seen in most tissues. Not
CC expressed in liver and peripheral blood leukocytes.
CC -1- SIMILARITY: Belongs to the GDAL / CD39 NTPase family.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC ENBL; AF039917; AAC39884.1; -.
CC ENBL; AF034840; AAC09236.2; -.
CC Genew; HGNC:3365; ENTPD3.
CC MIM; 603161; -.
CC InterPro; IPR000407; GDAL_CD39_NTPase.
CC Pfam; PF01150; GDAL_CD39; 1.
CC PROSITE; PS01238; GDAL_CD39_NTPase; 1.
CC Hydrolyase; Transmembrane; Glycoprotein; Calcium; Magnesium.
KW DOMAIN 1 22 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 23 43 POTENTIAL.
FT DOMAIN 44 485 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 486 506 POTENTIAL.
FT DOMAIN 507 529 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 81 81 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 149 149 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 402 402 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 454 454 R->G: INCREASE OF ACTIVITY.
FT MUTAGEN 67 67 R->A: LOSS OF ACTIVITY.
FT MUTAGEN 143 143 R->K: INCREASE OF ACTIVITY.
FT MUTAGEN 143 143 R->N: NO EFFECT.
FT MUTAGEN 146 146 R->P: INCREASE OF ATPASE ACTIVITY,
FT MUTAGEN 146 146 DECREASE OF ADPASE ACTIVITY.
FT MUTAGEN 146 146 R->T: INCREASE OF ACTIVITY.
FT MUTAGEN 182 182 E->D: COMPLETE LOSS OF ACTIVITY.
FT MUTAGEN 182 182 E->Q: COMPLETE LOSS OF ACTIVITY.
FT MUTAGEN 187 187 W->A: COMPLETE LOSS OF ACTIVITY.
```

```
FT MUTAGEN 191 191 N->A: LOSS OF ATPASE ACTIVITY, INCREASE
FT MUTAGEN 219 219 OF ADPASE ACTIVITY.
FT MUTAGEN 224 224 D->E: INCREASE OF ACTIVITY.
FT MUTAGEN 226 226 S->A: COMPLETE LOSS OF ACTIVITY.
FT MUTAGEN 459 459 Q->A: LOSS OF ACTIVITY.
FT MUTAGEN 459 459 W->A: INCREASE OF ACTIVITY, ESPECIALLY
THE ATP HYDROLYSIS.
FT CONFLICT 496 496 V -> A (IN REF. 3).
SQ SEQUENCE 529 AA; 59133 MW; 505D2PEC12978B88 CRC64;
```

Query Match 13.3%; Score 314.5; DB 1; Length 529;  
Best Local Similarity 25.3%; Pred. No. 2.7e-17;  
Matches 113; Conservative 76; Mismatches 166; Indels 91; Gaps 20;

```
QY 74 YGIMPDASTGTRRVHVFQTRPRTPTLTHETK-AVK-PGLSAYADVEXSAQGIREL 131
Db 57 YGIVADASSRTVYVQWPAKENTNGVVSQTRKCSYKSGISSYGNPDVPRAFEEC 116
QY 132 LDVAKQDIPDFFWKATPLVLKATAGLRLLPGEK--AQKLQKYKEVFKASPLVGDCC- 187
Db 117 MOKXKGVPSHLHSGTPIHLGATGMRLLRLQNETAANEVLESIOSYFKSQPF---DFR 172
QY 188 -VSIMNGTDEGVSAWITINFLTGS-----LKTPEGSSVGMLDIGGSGTQIAFLPR 236
Db 173 GAOIISGOEBGVYGTANYLMGNFLEKNLMHMMVHPHGVETTGALDLGGASTQISFVAG 232
QY 237 VEGTLAGSPPGVLTALRMFNRTYKLXSYSLGLGMSAR---LAILGVEGQPAKDGKEL 293
Db 233 EKMDLNTSD---IMOVSLXGYVYTLTYHSFOCYGRNEAKKFLMNL--LQNSPTKN--HL 285
QY 294 VSPCLSPSFKGEWEHAEV-----TYRVSGQKAAASLHELCAARVS 333
Db 286 TNPCTPRDYISLFTWGHVFDLSCTVDQRPESYNPDVITFEGTDPG-----LCKEKVA 339
QY 334 EVLQWRVHRTBEVKHVD-----FYAFSYVYDLAAGVGLDAEKGSLLVYGDPEIA 383
Db 340 SIIDFKACHDQETCSFDGVYQPKIKGPVAFAGFYTASALNL-----SGSFSLDTNNS 394
QY 384 AKYVCRLETC-POSSP-----FSCMDLTYV-SLLDEFGPPRS--KYLKLTIRKIDN 431
Db 395 TWNFCSQWMSQPLPLPKFDEVYARSYCFSANITYHLFVNGYKFTLETWPIIHFEKEVGN 454
QY 432 VETSWALGATFHVYDLSLRQ--KSP 454
Db 455 SSIAWSLG---YMLSLTNQIPASBP 476
```

Search completed: June 8, 2004, 09:48:49  
Job time : 20 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: June 8, 2004, 09:48:15 ; Search time 46 Seconds  
(without alignments)  
3127.746 Million cell updates/sec

Title: US-09-905-589a-2  
Perfect score: 2364  
Sequence: 1 MRKISNHSIRVAKVAVPLG.....ALGAIHFYIDSLNRKSPAS 456

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues  
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: sp\_archaea: \*  
2: sp\_bacteria: \*  
3: sp\_fungi: \*  
4: sp\_human: \*  
5: sp\_invertebrate: \*  
6: sp\_mammal: \*  
7: sp\_mhc: \*  
8: sp\_organelle: \*  
9: sp\_phage: \*  
10: sp\_plant: \*  
11: sp\_rodent: \*  
12: sp\_virus: \*  
13: sp\_vertebrate: \*  
14: sp\_unclassified: \*  
15: sp\_virus: \*  
16: sp\_bacteriopl: \*  
17: sp\_archaeap: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2361	99.9	483	4	Q8TAS7
2	2361	99.9	484	4	Q7Z5B5
3	2361	99.9	503	4	Q8N3H3
4	1229.5	52.0	278	11	Q8CH23
5	1007	42.0	427	11	Q8CD29
6	1000	42.3	427	11	Q8BR23
7	992	42.0	428	4	Q8ERX0
8	957	38.5	224	11	Q8CDB6
9	909	36.5	407	4	Q8WUB3
10	621	26.3	461	5	Q7Z6Z8
11	621	26.3	464	5	Q8YU18
12	590.5	25.0	479	5	Q8XU84
13	531.5	22.5	522	3	Q8HEX6
14	528	22.2	489	10	Q8H7L6
15	525.5	22.2	467	10	Q8SPW7
16	509	21.5	455	10	Q84UE0

17	508	21.5	556	3	Q9UT35	Q9UT35 schizosach
18	506	21.4	462	10	Q9XFC9	Q9XFC9 dolichos bi
19	505.5	21.4	467	10	Q84UD8	Q84UD8 medicago tr
20	504.5	21.3	467	10	Q8RV76	Q8RV76 pisum sativ
21	501.5	21.2	447	10	Q9FU11	Q9FU11 pisum sativ
22	501	21.2	455	10	Q9SPW6	Q9SPW6 medicago sa
23	489.5	20.7	471	10	Q8SGQ2	Q8SGQ2 arabidopsis
24	489	20.7	599	3	Q8TGH6	Q8TGH6 candida alb
25	488.5	20.7	472	10	Q9SPM5	Q9SPM5 arabidopsis
26	486.5	20.6	462	10	Q84UE2	Q84UE2 medicago tr
27	485.5	20.5	472	10	Q9M7B3	Q9M7B3 arabidopsis
28	485	20.5	468	10	Q9FVC2	Q9FVC2 glycine soj
29	484.5	20.5	472	10	Q8L704	Q8L704 arabidopsis
30	484	20.5	463	10	Q8FVC3	Q8FVC3 glycine soj
31	483.5	20.5	558	3	Q8TGS8	Q8TGS8 aspergillus
32	481	20.3	467	10	Q84UE1	Q84UE1 medicago tr
33	463	19.6	456	10	Q9SPM8	Q9SPM8 lotus japon
34	462	19.5	455	10	Q84UD9	Q84UD9 medicago tr
35	454.5	19.2	454	10	Q84UE3	Q84UE3 medicago tr
36	447.5	18.9	466	10	Q9AUI5	Q9AUI5 medicago tr
37	442.5	18.7	455	10	Q9FEX6	Q9FEX6 pisum sativ
38	442	18.7	556	3	Q9C2M0	Q9C2M0 neurospora
39	441.5	18.7	455	10	Q84L88	Q84L88 glycine max
40	439.5	18.6	455	10	Q8GTB1	Q8GTB1 pisum sativ
41	434.5	18.4	447	10	Q9SLV4	Q9SLV4 pisum sativ
42	430.5	18.2	455	10	Q8RV10	Q8RV10 pisum sativ
43	425.5	18.0	407	10	Q8AVN8	Q8AVN8 pisum sativ
44	415.5	17.6	473	10	Q8RV78	Q8RV78 pisum sativ
45	396.5	16.8	455	10	Q8RV79	Q8RV79 pisum sativ

## ALIGNMENTS

RESULT 1	ID	Q8TAS7	PRELIMINARY;	PRT;	483 AA.
AC	Q8TAS7				
DT	01-JUN-2002 (TrEMBLrel. 21, Created)				
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)				
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)				
DE	Ectonucleoside triphosphate diphosphohydrolase 6 (putative DE function).				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Placenta;				
RA	Strausberg R.;				
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.				
DR	EMBL: BC025980; AAH25980.1; -				
DR	GO: GO:0016787; F-1-hydrolase activity; IEA.				
DR	InterPro: IPR000407; GDAI_CD39_NTPase.				
DR	Pfam: PFO150; GDAI_CD39; 1.				
KW	Hydrolase.				
SC	SEQUENCE 483 AA; 5119 MW; A8E0E5035BCDC8F CRC64;				
Query Match	99.9%;	Score 2361;	DB 4;	Length 483;	
Best Local Similarity	99.8%;	Pred. NO. 6.8e-196;			
Matches 455;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	MRKISNHSIRVAKVAVPLGCVGFIVYVYIKWRAATATQAFPSITRAAPARMCQAH	60		
DB	28	MRKISNHSIRVAKVAVPLGCVGFIVYVYIKWRAATATQAFPSITRAAPARMCQAH	87		
QY	61	SPRTGADGHEVYVYGMFAGSGTGVVHFOFTRPREPTLTHFFKVGKGLAVYDD	120		
DB	88	SPRTGADGHEVYVYGMFAGSGTGVVHFOFTRPREPTLTHFFKVGKGLAVYDD	147		
QY	121	VEKSAQIGRELLDVAKQDIPDFWKATPLVLKATAGRLLPGEKQKLLQXKVEFKSP	180		

```
Db 148 VKSAQGIKRELLDVAQKQDIPDFWKATPLVLKATAGRLLLPGEKAQKLLQKVKEVFKASP 207
QY 181 FLVGDDCVSIMNGTDEGVSAWITINFLTGS�KTPGGSSVGMGLDLGGSTQIAFLPRVEGT 240
Db 208 FLVGDDCVSIMNGTDEGVSAWITINFLTGS�KTPGGSSVGMGLDLGGSTQIAFLPRVEGT 267
QY 241 LQASPPGYLTALRMFNRTYKLYSYLGLGMSARLAILGGVEGQPAKDGKELVSPCLSP 300
Db 268 LQASPPGYLTALRMFNRTYKLYSYLGLGMSARLAILGGVEGQPAKDGKELVSPCLSP 327
QY 301 SFKGEWEHAETVYRVSGQKAAASLHELCAARVSEVLQNRVHRTTEVKHVDVFAFSYYDL 360
Db 328 SFKGEWEHAETVYRVSGQKAAASLHELCAARVSEVLQNRVHRTTEVKHVDVFAFSYYDL 387
QY 361 AAGVGLIDAEKGSLLVVGDFEIAAKVYCRLETQPOSSPSCMDLTYVSLLLQEFPPRS 420
Db 388 AAGVGLIDAEKGSLLVVGDFEIAAKVYCRLETQPOSSPSCMDLTYVSLLLQEFPPRS 447
QY 421 KVLKLRKIDNVETSWALGAIFHYIDSLNRQKSPAS 456
Db 448 KVLKLRKIDNVETSWALGAIFHYIDSLNRQKSPAS 483

RESULT 2
Q725B5 PRELIMINARY; PRT; 484 AA.
AC Q725B5;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DE CD39L2 nucleotidease.
GN ENTPD6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Ivanenkov V.V., Murphy-Piedmonte D.M., Kirley T.L.;
RT "Bacterial Expression, Characterization, and Disulfide Bond
RT Determination of Soluble Human NTPase6 (CD39L2) Nucleotidease:
RT Implications for Structure and Function.";
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR ENBL; AY227581; AAP92131.1; -.
SQ SEQUENCE 484 AA; 53274 MW; D00A5D2915DF36CE CRC64;

Query Match 99.9%; Score 2361; DB 4; Length 484;
Best Local Similarity 99.8%; Pred. No. 6.9e-196;
Matches 455; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKISHGSLRVAKVAYPLGCVGVFIYVAYIKWHRATATQAFSITRAAPGARWGOQAH 60
Db 29 MRKISHGSLRVAKVAYPLGCVGVFIYVAYIKWHRATATQAFSITRAAPGARWGOQAH 88
QY 61 SPLGTAADGHEVYGMFMDAGSTGTRVHVFPQTRPREPTLTHETFAKPKGLSAYADD 120
Db 89 SPLGTAADGHEVYGMFMDAGSTGTRVHVFPQTRPREPTLTHETFAKPKGLSAYADD 148
QY 121 VKSAQGIKRELLDVAQKQDIPDFWKATPLVLKATAGRLLLPGEKAQKLLQKVKEVFKASP 180
Db 149 VKSAQGIKRELLDVAQKQDIPDFWKATPLVLKATAGRLLLPGEKAQKLLQKVKEVFKASP 208
QY 181 FLVGDDCVSIMNGTDEGVSAWITINFLTGS�KTPGGSSVGMGLDLGGSTQIAFLPRVEGT 240
Db 209 FLVGDDCVSIMNGTDEGVSAWITINFLTGS�KTPGGSSVGMGLDLGGSTQIAFLPRVEGT 269
QY 241 LQASPPGYLTALRMFNRTYKLYSYLGLGMSARLAILGGVEGQPAKDGKELVSPCLSP 300
Db 269 LQASPPGYLTALRMFNRTYKLYSYLGLGMSARLAILGGVEGQPAKDGKELVSPCLSP 328
QY 301 SFKGEWEHAETVYRVSGQKAAASLHELCAARVSEVLQNRVHRTTEVKHVDVFAFSYYDL 360
```

```
Db 329 SFKGEWEHAETVYRVSGQKAAASLHELCAARVSEVLQNRVHRTTEVKHVDVFAFSYYDL 388
QY 361 AAGVGLIDAEKGSLLVVGDFEIAAKVYCRLETQPOSSPSCMDLTYVSLLLQEFPPRS 420
Db 389 AAGVGLIDAEKGSLLVVGDFEIAAKVYCRLETQPOSSPSCMDLTYVSLLLQEFPPRS 448
QY 421 KVLKLRKIDNVETSWALGAIFHYIDSLNRQKSPAS 456
Db 449 KVLKLRKIDNVETSWALGAIFHYIDSLNRQKSPAS 484

RESULT 3
Q8N3H3 PRELIMINARY; PRT; 503 AA.
AC Q8N3H3;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Hypothetical protein (fragment).
GN DKFZP761J1915.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Amvgdala;
RA Ansgorge W., Wirkner U., Mewes H.W., Weill B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR ENBL; AL834158; CAD38864.1; -.
DR InterPro; IPR000407; GDAI_CD39_NTPase.
DR Pfam; PF01150; GDAI_CD39; 1.
DR KX Hypothetical protein.
FT NON TER
SQ SEQUENCE 503 AA; 54763 MW; 163933P9139D89F CRC64;

Query Match 99.8%; Score 2361; DB 4; Length 503;
Best Local Similarity 99.8%; Pred. No. 7.3e-196;
Matches 455; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKISHGSLRVAKVAYPLGCVGVFIYVAYIKWHRATATQAFSITRAAPGARWGOQAH 60
Db 48 MRKISHGSLRVAKVAYPLGCVGVFIYVAYIKWHRATATQAFSITRAAPGARWGOQAH 107
QY 61 SPLGTAADGHEVYGMFMDAGSTGTRVHVFPQTRPREPTLTHETFAKPKGLSAYADD 120
Db 108 SPLGTAADGHEVYGMFMDAGSTGTRVHVFPQTRPREPTLTHETFAKPKGLSAYADD 167
QY 121 VKSAQGIKRELLDVAQKQDIPDFWKATPLVLKATAGRLLLPGEKAQKLLQKVKEVFKASP 180
Db 168 VKSAQGIKRELLDVAQKQDIPDFWKATPLVLKATAGRLLLPGEKAQKLLQKVKEVFKASP 227
QY 181 FLVGDDCVSIMNGTDEGVSAWITINFLTGS�KTPGGSSVGMGLDLGGSTQIAFLPRVEGT 240
Db 228 FLVGDDCVSIMNGTDEGVSAWITINFLTGS�KTPGGSSVGMGLDLGGSTQIAFLPRVEGT 287
QY 241 LQASPPGYLTALRMFNRTYKLYSYLGLGMSARLAILGGVEGQPAKDGKELVSPCLSP 300
Db 288 LQASPPGYLTALRMFNRTYKLYSYLGLGMSARLAILGGVEGQPAKDGKELVSPCLSP 347
QY 301 SFKGEWEHAETVYRVSGQKAAASLHELCAARVSEVLQNRVHRTTEVKHVDVFAFSYYDL 360
Db 348 SFKGEWEHAETVYRVSGQKAAASLHELCAARVSEVLQNRVHRTTEVKHVDVFAFSYYDL 407
QY 361 AAGVGLIDAEKGSLLVVGDFEIAAKVYCRLETQPOSSPSCMDLTYVSLLLQEFPPRS 420
Db 408 AAGVGLIDAEKGSLLVVGDFEIAAKVYCRLETQPOSSPSCMDLTYVSLLLQEFPPRS 467
QY 421 KVLKLRKIDNVETSWALGAIFHYIDSLNRQKSPAS 456
Db 468 KVLKLRKIDNVETSWALGAIFHYIDSLNRQKSPAS 503
```

```

RESULT 4
Q8CH23 PRELIMINARY; PRT; 278 AA.
AC Q8CH23;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Similar to ectonucleoside triphosphate diphosphohydrolase 6.
GN ENTPD6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
RA Strauberg R.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC038126; AAI38126.1; -.
DR MGD: MGI:1202295; Entpd6.
DR GO: GO:0016787; F:hydrolase activity; IEA.
DR InterPro: IPR00407; GDAI_CD39_NTPase.
DR Pfam: PF01150; GDAI_CD39; 1.
DR Hydrolase.
SQ SEQUENCE 278 AA; 30130 MW; 75A92DD1AC76297F CRC64;

Query Match 52.0%; Score 1229.5; DB 11; Length 278;
Best Local Similarity 85.6%; Pred. No. 4e-98;
Matches 237; Conservative 17; Mismatches 18; Indels 5; Gaps 1;

QY 1 MRKISNGSLRAVAVAYPLGCVGFYVAVYIKHRAATATQAFSITRAPAGARWGOAH 60
DB 1 MRKIPNNGTLMRTVAAPLGLCVGLFTYVAYIKHRAAQAFTTLAGASGARWGOAH 60

QY 61 SPLGTADGHEVYFYGIMPDAGSTGTRVHVFQFTRPREPTTLTHTFKAVKPGISAVAD 120
DB 61 SSPSARGHEVYFYGIMPDAGSTGTRVHVFQFTRPREPTTLTHTFKAVKPGISAVAD 120

QY 121 VERAAGIRELIDVAKODIPDFPKATPLVKAATAGRLGPKAKOKLQKVEVRKASP 180
DB 121 VERAAGIOELINVAKHQIIPDFPKATPLVKAATAGRLGPKAKOKLQKVEVRKASP 180

QY 121 FLVGDGVSINMGDEGSAMITINFLTGLSKTPGSSVGMLDLGGSGTOIAELPRVEGT 240
DB 121 FLVGDGVSINMGDEGSAMITINFLTGLSKTPGSSVGMLDLGGSGTOITLPRVEGT 240

QY 181 FLVGDGVSINMGDEGSAMITINFLTGLSKTPGSSVGMLDLGGSGTOIAELPRVEGT 240
DB 181 FLVGDGVSINMGDEGSAMITINFLTGLSKTPGSSVGMLDLGGSGTOITLPRVEGT 240

QY 241 LQASPPGYLTALRMFRTTYKLYSYGLGLMSARLA 277
DB 241 LQASPPGYLTALRMFRTTYKLYSYR-----VCSRLA 272

RESULT 5
Q8CD29 PRELIMINARY; PRT; 427 AA.
AC Q8CD29;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Ectonucleoside triphosphate diphosphohydrolase 5.
GN ENTPD5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium;
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).

```

```

DR EMBL: AK031581; BAC27461.1; -.
DR MGD: MGI:1321385; Entpd5.
DR InterPro: IPR00407; GDAI_CD39_NTPase.
DR Pfam: PF01150; GDAI_CD39; 1.
SQ SEQUENCE 427 AA; 47101 MW; 6E3773C842B58477 CRC64;

Query Match 42.6%; Score 1007; DB 11; Length 427;
Best Local Similarity 50.0%; Pred. No. 1.6e-78;
Matches 213; Conservative 62; Mismatches 133; Indels 18; Gaps 6;

QY 37 ATATQAFSITRAPAGA-----RMGO-----QAHSLGTADGHEVYFYGIMPDAGSTG 84
DB 2 ATSWGAVMLLIACGSGVFYREQQTWEQVFLSMCINVSAG---TFYIGMFDAGSTG 58

QY 85 TRVAVFQRT-RPREPTPLTHTETKAVKPGISAYADVEKSAQIGRELIDVAKODIPDF 143
DB 59 TRIHVYTFVQKTAGQLPFLBGEIIPDSVKGSAFVDPKQAEIVQELLEVAKOSIPRSH 118

QY 144 WKATPLVTKAATAGRLGPKAKOKLQKVEVRKASPFLVGDGVSINMGDEGSAMIT 203
DB 119 WERTVVKATRGRLPLFEQKQALLVEBEIFKNSPFLVDSVSTMDGSYEGILAMVT 178

QY 204 INFLTGLSKTPGSSVGMLDLGGSGTOIAELPRVEGTLOASPPGYLTALRMFRTTYKLYS 263
DB 179 VNFLTGQLHGRGOETVGTLDLGGASTGITFLPQFEKTELEKTPRGYLTSEFMFNSTFKLYT 238

QY 264 VSYLGLGMSARIALTGVGEGPAKDKELVSPCLSPFKEWEMHAEVTVSGOKAAS 323
DB 239 HSYLDFGKAAPLALTLGLALENK-GTGDHTRFSACLPRMLEMWFPGVYKIQYGNGSGEM 297

QY 324 LHELCAAVSEVLTQNRVHRTVEVKHVDFAFSYYDDLAAGVGLDAEKGSLVVDPEFLA 383
DB 296 GEPFCYAEVLARVQCKHQPEVEGSAFYASYYDDRAADTHLIDYEGVGLKVEDPERK 357

QY 384 AKYVKTLETQDQSPFSCMDLITYSLLOE-FGEPKAKLTKRKIDNVTSMALGIF 442
DB 358 AEVCDNIGSPSSGSPFLCMDLITYTALLKQGFADGTLQTLTKVNNIETGMALGTF 417

QY 443 HYDSL 448
DB 418 HLQSL 423

RESULT 6
Q8BR23 PRELIMINARY; PRT; 427 AA.
AC Q8BR23;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Ectonucleoside triphosphate diphosphohydrolase 5.
GN ENTPD5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Brain;
RA MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium;
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL: AK045828; BAC32507.1; -.
DR MGD: MGI:1321385; Entpd5.
DR InterPro: IPR00407; GDAI_CD39_NTPase.
DR Pfam: PF01150; GDAI_CD39; 1.
SQ SEQUENCE 427 AA; 47123 MW; 87BF2CC1CC1FCB9 CRC64;

Query Match 42.3%; Score 1000; DB 11; Length 427;
Best Local Similarity 49.8%; Pred. No. 6.5e-78;

```



AC Q9WUB3; 01-MAR-2002 (TRENBLREL. 20, Created)  
 DT 01-MAR-2002 (TRENBLREL. 20, Last sequence update)  
 DT 01-MAR-2002 (TRENBLREL. 20, Last annotation update)  
 DE Similar to econucleoside triphosphate diphosphohydrolase 5.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Tissue=Colon;  
 RA Struysberg R.;  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC020966; AAH20966.1; -  
 DR GO; GO:0016787; F:hydrolase activity; ISA.  
 DR InterPro; IPR000407; GDA1\_CD39\_NTPase.  
 DR Pfam; PF01150; GDA1\_CD39; 1.  
 KW Hydrolase.  
 SQ SEQUENCE 407 AA; 45336 MW; D92A5F7DC9EC9E5B CRC64;  
 Query Match 38.5%; Score 909; DB 4; Length 407;  
 Best Local Similarity 51.8%; Pred. No. 4.7e-70;  
 Matches 189; Conservative 51; Mismatches 119; Indels 6; Gaps 4;  
 QY 62 PLGTADGHEVYFYMFGASTGTRVHVFQFT-RRPRRTPLTHTETFAVKGISAYADD 120  
 DB 40 PINVA---STYGLMFDAGSTGTRHVTYQKPGQLTLEGEVFDSPVPSAFVDQ 96  
 QY 121 VEKSAQIRELLDVAKODIPDFFWKATPLVKAINGLRLPGEKAKLQYKVEFVAFSP 180  
 DB 97 PRQGEATYQGLEVAKODIPRSHMKKTPLVKAINGLRLPGEKAKLLEFVKEIFRSP 156  
 QY 161 FLVGGDCISIMNGTDEGVSAWTTIFLGLSKTPEGSSVGMLDGGSTQAFPRVBT 240  
 DB 157 FLVPGGSVIMDGSDEGLAWTVFLTGQHGHRQETVGLDGGASTQTFPQPEKT 216  
 QY 241 IQASPPGYLTALRMNRTYKLYSYLGLGMSARLALIGVEGQPAKGEIVSPCLSP 300  
 DB 217 LEQTRGVLTSPFEMNSYKLYTHSYLFGKLAALATLGLMTE-GRDGHTRFSACIPR 275  
 QY 301 SFKGEWHEAEVTVRVSGQKAAASLHETCAARYSEVLQNRKVRTEVKAADVAFSYIDL 360  
 DB 276 WLEAWMITGQVYQYQGNQGEVGFPEYAEVLRVVRKQLQPEBVGSGSYAFASYIDR 335  
 QY 361 AAGVGLIDAEKGSVWGFELIAKYVCRLETQOSPFSCMDLYVSLLOE-FGPPR 419  
 DB 336 AVDTMIDYERKGIKVEDFERKAREVCDNLENFTSSGSPFLCMLSYITALLKQGFAD 395  
 QY 420 SKVLK 424  
 DB 396 STVLQ 400  
 RESULT 10  
 OT6268 PRELIMINARY; PRT; 461 AA.  
 AC 076268;  
 DT 01-NOV-1998 (TRENBLREL. 08, Created)  
 DT 01-NOV-1998 (TRENBLREL. 08, Last sequence update)  
 DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)  
 DE NTPase protein (JD116412).  
 GN NTPASE OR CG3059.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NC NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=2019606; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., McKee G.L.G.,  
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
 RA Baller R.M., Basu A., Bakendale U., Bayraktiroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhanderi D., Boshakov S.,  
 RA Borikova D., Borchan M.R., Bouck J., Brokstein P., Brothman P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart M.M., Glaeser K.,  
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lascko P., Lei Y., Levitsky A.A., Li U., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Marklov G., Malishina N.V., Moberly C., Morris D., Mostrel A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Rainert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Sytkas R., Teclor C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of *Drosophila melanogaster*."  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98341119; PubMed=9676430;  
 RA Chadwick B.P., Fritschau A.M.,  
 RT "The CD39-like gene family: Identification of three new human members  
 RT (CD39L2, CD39L3, and CD39L4), their murine homologues, and a member of  
 RT the gene family from *Drosophila melanogaster*."  
 RL Genomics 50:357-367(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nuno J., Pacle J., Paragay V., Park S., Phouanavong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AEO03581; AAF51182.1; -  
 DR EMBL; AF041048; AAC93133.1; -  
 DR EMBL; AY061134; AAL28682.1; -  
 DR FLYBase; FBgn0024947; NTPase.  
 DR InterPro; IPR000407; GDA1\_CD39\_NTPase.  
 DR Pfam; PF01150; GDA1\_CD39; 1.  
 SQ SEQUENCE 461 AA; 50845 MW; 27D00321F91A9DD1 CRC64;  
 Query Match 26.3%; Score 621; DB 5; Length 461;  
 Best Local Similarity 39.9%; Pred. No. 5.6e-45;  
 Matches 158; Conservative 60; Mismatches 142; Indels 35; Gaps 13;  
 QY 71 EVFYGLMFDAGSTGTRVHVFQFT-RRPRRTPLTHTETFAVKGISAYADDVEKSAQGIR 129  
 DB 76 KQYLAIIIDAGSTGTRVHVFQFT-RRPRRTPLTHTETFAVKGISAYADDVEKSAQGIR 135  
 QY 130 ELIDVAKODIPDFFWKATPLVKAINGLRLPGEKAKLQYKVEFVAFSPVLDGDCVS 189  
 DB 136 LLIDDEARAFIPKHEWSTPLVKAINGLRLPGEKAKLQYKVEFVAFSPVLDGDCVS 195





```
Db 242 GSVQWYALSEKADGAPPAEGBSDSYVKELLLKGTYYLVVHSYLRGLLAARAEILKA 301
QY 282 VEGQPAKDGKELVSPCLSPSPKGEWEHAEVTVYRVSQKAAASLHELCAARYSEVLQNRVH 341
Db 302 GEGNDYRN-----CMLEGHGQYRGDDIFEASGLSSGASYSKCRVAV-----RAL 348
QY 342 RTEE--VRHV-----DFYAFSYTYDAAAGVGLIDAEKGGSLV-VGDPE 381
Db 349 KYDEPACTHMKCTFGGVWNGGGGQKNLFVASFDFDRAAEAGFVNPKAPFAKVPKPSDFE 408
QY 382 IAAKYVCR-----TLETQOSS-----PFSCMDLTY-VSLLLOEGRFPKSKVLKTRKI-- 429
Db 409 EAARRVCKLVNDKAQATYPDVSEENVLYCLMDLVYQYTLVVDGFGVDYQDITLVKKVPY 468
QY 430 --DNVETSWALGAIFHYIDS 447
Db 469 SNSFVEAAWPLGSAIEVASS 488
```

```
RESULT 15
Q9SPM7
ID Q9SPM7 PRELIMINARY; PRT; 467 AA.
AC Q9SPM7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Apyrase.
OS Dolichos biflorus (Horse gram).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Dolichos.
OX NCBI_TaxID=3840;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=99444909; PubMed=10517321;
RA Roberts N.J., Brigham J., Wu B., Murphy J.B., Volpin H.,
RA Phillips D.A., Etzler M.E.;
RT "A Nod factor-binding lectin is a member of a distinct class of
RT apyrases that may be unique to the legumes.";
RL Mol. Gen. Genet. 262:261-267(1999).
DR EMBL; AF156781; AAF00610.1; -.
DR InterPro; IPR000407; GDAL_CD39_NTPase.
DR Pfam; PF01150; GDAL_CD39; 1.
DR PROSITE; PS01238; GDAL_CD39_NTPASE; 1.
SQ SEQUENCE 467 AA; 51164 MW; 8FCC200AA60D7376 CRC64;
```

```
Query Match 22.2%; Score 525.5; DB 10; Length 467;
Best Local Similarity 31.7%; Pred. No. 1.1e-36;
Matches 145; Conservative 85; Mismatches 162; Indels 65; Gaps 17;
QY 37 ATATQAFSITRAAPGARWQQAHSPLGTAADGHEV-----FYGIMFDAGSTGRVHV 90
Db 21 ATATASFSL--HGKGFK-----HRKFSSDDNNYSIEETINESYAVIFDAGSTGSRVHV 73
QY 91 QFTRPREPTFTLTH--ETPKAVKPGLSAYADVDEKSAQCIRELLDVAKODIPDFWKATP 149
Db 74 RFNQ-QLDLRLIGHDLLELVKTKPGLSAYAPENPEAAESLVLPLLEAEAVIPQELHPTP 132
QY 149 LVLKATAGLRLPGEKAQKLQKKEVP-KASPFVVGDDCVSIMNGTDEGVSAWITINFL 207
Db 133 VKVGATAGLRQLLEGASNRILQAVSMLKRTLKVEGDVAVSLSGNQEGAYQWVTIINYL 192
QY 208 TGS�KTPGGSSVGMLDLGGGSGTQIAFLPRVEGTQAS--PPG---YLTALRMFNFTKLY 262
Db 193 LGNLGHYSKTVAVVDLGGSVQMAVISEDAAKAPQVDPGVESYITEMFLRGKKYLY 252
QY 263 SYSYGLGLMSARLAILGGVEGPAKDGKELVSPCLSPSKGEWEHAEVTVYRVSQKAAA 322
Db 253 VHSYLRGLLAARAEVL-----KVRSDSENPCILSGFDGTYTYGGVQYKATAPPSSG 304
QY 323 SLHELCAARYSEVLQNRVHRTVEVKHV-----DFVAFSYTYDLAAGVGLI 367
```

```
Db 305 SFSK-QQNVVLEAL--HVNATCSYKDCPTGGIWNQGGGAGENNPFVASFVEVADGAFV 361
QY 368 DAKEGGSLV-VGDFEIAAKYVCRKTLETQPOSS-----PFSCMDLTY-VSLLLOEF 415
Db 362 DPNDANAIVRPVDFEDAAKVACST-ELKDLKSVFPRVKDGDVPYICLDLVYQYTLVVDGF 420
QY 416 GFPRSKVLKTRKIDN-----VETSWALGAIFHYIDSL 448
Db 421 GIDPQOEITLVROIQYQDSLSVEAAWPLGSAIEAIISSL 457
```

Search completed: June 8, 2004, 09:51:33  
Job time : 49 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: June 8, 2004, 09:48:14 ; Search time 59 Seconds  
(without alignments)  
2183.757 Million cell updates/sec

Title: US-09-905-589A-2

Perfect score: 2364  
Sequence: 1 MKRISMGSLRVAKVAVPLG.....ALGATPHYIDSLNRKSPAS 456

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseq1980s:\*  
2: geneseq1990s:\*  
3: geneseq2000s:\*  
4: geneseq2001s:\*  
5: geneseq2002s:\*  
6: geneseq2003as:\*  
7: geneseq2003bs:\*  
8: geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2364	100.0	456	5	AAE19881 Human CD3
2	2364	100.0	484	4	AAE19881 Human CD3
3	2361	99.9	463	5	ABO6124 Human NS
4	2361	99.9	467	5	ABJ04657 Protein o
5	2351	99.5	456	4	AAm93929 Human pol
6	2152	91.0	450	7	ADCI4220 Human enz
7	2135	90.3	462	4	AAU30882 Novel hum
8	2116	89.5	446	5	ABJ04658 Protein o
9	2003	84.7	471	4	AAE19881 Human CD3
10	999	42.3	428	4	AAE19881 Human CD3
11	999	42.3	428	4	AAE19881 Human CD3
12	999	42.3	428	4	AAE19881 Human CD3
13	999	42.3	428	5	AAE19881 Human CD3
14	996	42.1	428	3	AAE19881 Human CD3
15	996	42.1	428	3	AAE19881 Human CD3
16	909	38.5	405	3	AAE19881 Human CD3
17	909	38.5	405	3	AAE19881 Human CD3
18	904	38.2	465	5	AAE19881 Human CD3
19	823	34.8	330	3	AAE19881 Human CD3
20	621	26.3	461	4	AAE19881 Human CD3
21	621	26.3	461	4	AAE19881 Human CD3
22	596	25.2	141	4	AAU30879 Novel hum
23	596	25.2	144	4	AAU30881 Novel hum
24	518.5	21.9	467	2	AAW85687 DBX oligo
25	506	21.4	462	2	AAW85684 NBP46 roo

26	506	21.4	462	5	AAU78818 Dolichos
27	501	21.2	486	2	AAW85685 NBP46 roo
28	501	21.2	486	5	AAU78819 Lotus jap
29	488.5	20.7	472	6	ABP81286 Arabidops
30	486	20.6	457	7	AAE19881 Human CD3
31	477	20.2	407	7	AAE19881 Human CD3
32	463	19.6	496	5	AAU78820 Medicago
33	454	19.2	496	2	AAW85686 NBP46 roo
34	441.5	18.7	455	2	AAE19881 Human CD3
35	414	17.5	105	4	ABP69389 Human pol
36	371	15.7	139	4	AAU30880 Novel hum
37	314.5	13.3	529	5	AAE19881 Human CD3
38	314.5	13.3	529	5	AAU76973 Human CD3
39	304.5	12.9	604	5	ABP64710 Human pro
40	304.5	12.9	604	5	AAE19881 Human CD3
41	304.5	12.9	604	7	AAU09133 Novel hum
42	287.5	12.2	478	3	AAU70914 Human sol
43	287.5	12.2	478	3	AAU70891 Protein e
44	287	12.1	476	3	AAU70912 Human CD3
45	287	12.1	476	3	AAU70889 Protein e

## ALIGNMENTS

RESULT 1	AAE19881	standard; protein; 456 AA.
ID	AAE19881	standard; protein; 456 AA.
AC	AAE19881	
XX		
DT	18-JUN-2002	(first entry)
XX		
DE	Human CD39L2 protein.	
XX		
XX	Human; CD-39-like protein; CD39L2 protein; therapy; immune deficiency; autoimmune disorder; multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis; autoimmune thyroiditis; allergic reaction; asthma; insulin dependent diabetes mellitus; periodontal disease; osteoporosis; osteoarthritis; wound healing; tissue repair; Alzheimer's disease; ulcer; Parkinson's disease; amyotrophic lateral sclerosis; Huntington's disease; nervous system disease; nerve injury; ischaemia-reperfusion injury; endotoxin lethality; arthritis; nephritis; inflammatory bowel disease; Crohn's disease; vitiligo; antibacterial; antifungal; neuroprotective; dermatological; immunosuppressive; vulnery; noctropic; anticonvulsant; antiinflammatory; nephrotropic; gastrointestinal; vasotropic.	
XX		
OS	Homo sapiens.	
XX		
PN	US6350447-B1.	
XX		
PD	26-FEB-2002.	
XX		
PF	29-JAN-1999; 99US-00240639.	
XX		
PR	29-JAN-1999; 99US-00240639.	
XX		
PA	(HYSE-) HYSEQ INC.	
PI	Chadwick BP, Friesch A,	
XX		
DR	WPI; 2002-215262/27.	
XX		
DR	N-PDB; AAD1693.	
XX		
PT	An isolated polypeptide with phosphohydrolase activity, designated CD39L2, useful to identify other proteins with which binding occurs or identify inhibitors and for treatment of, e.g., Alzheimer's, multiple sclerosis and osteoporosis.	
PT		
PS	Claim 1, Fig 4, 10pp: English.	
XX		
CC	The present invention relates to novel proteins with phosphohydrolase activity, designated CD-39-like (CD39L) proteins and polynucleotides	

CC encoding such proteins, CD39L proteins are useful to treat infectious  
 CC diseases caused by viral, bacterial, fungal or other infection that may  
 CC be treatable with CD39L. They are useful in the treatment of various  
 CC immune deficiencies and disorders, autoimmune disorders such as multiple  
 CC sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune  
 CC thyroiditis and insulin dependent diabetes mellitus, allergic reactions  
 CC and conditions such as asthma and other respiratory problems, periodontal  
 CC disease, osteoporosis, osteoarthritis and other tooth repair processes.  
 CC They may have utility in compositions used for bone, cartilage, tendon,  
 CC ligament and/or nerve tissue growth or regeneration as well as for wound  
 CC healing and tissue repair and replacement and in the treatment of burns,  
 CC incisions and ulcers. CD39L proteins may also be useful for proliferation  
 CC of neural cells and for regeneration of nerve and brain tissue, i.e. for  
 CC the treatment of central nervous system diseases such as Alzheimer's  
 CC disease, Parkinson's disease, amyotrophic lateral sclerosis, Huntington's  
 CC disease, peripheral nervous system diseases peripheral nerve injuries,  
 CC peripheral neuropathy and localised neuropathies. They are also used to  
 CC treat mechanical and traumatic disorders which involve degeneration,  
 CC death or trauma to neural cells or nerve tissue. CD39L proteins of the  
 CC invention are also useful to promote better or faster closure of non-  
 CC healing wounds, including pressure ulcers, ulcers associated with  
 CC vascular insufficiency and surgical and traumatic wounds. They also  
 CC exhibit anti-inflammatory activity and may be used to treat inflammatory  
 CC conditions including chronic or acute conditions), including ischaemia-  
 CC reperfusion injury, endotoxin lethality, arthritis, nephritis, cytokine  
 CC or chemokine-induced lung injury, inflammatory bowel disease or Crohn's  
 CC disease. The present sequence is human CD39L2 protein  
 XX  
 SQ Sequence 456 AA;

Query Match 100.0%; Score 2364; DB 5; Length 456;  
 Best Local Similarity 100.0%; Pred. No. 5.4e-216;  
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKISNHGSLRVAKVAYPLGLCVGFIYVAYIKWHRATATQAFSTRAAPGAWGQQA 60  
 DB 1 MRKISNHGSLRVAKVAYPLGLCVGFIYVAYIKWHRATATQAFSTRAAPGAWGQQA 60  
 QY 61 SPLGTADGHEVYIGMFDAGSGTGRVHVQFTRPPRPTTLTHETFKAVKPGLSAYADD 120  
 DB 61 SPLGTADGHEVYIGMFDAGSGTGRVHVQFTRPPRPTTLTHETFKAVKPGLSAYADD 120  
 QY 121 VESAGQIRELLDVAKQDIPDFWKATPLVLKATAGRLLPGEKAKLQKVEFKASP 180  
 DB 121 VESAGQIRELLDVAKQDIPDFWKATPLVLKATAGRLLPGEKAKLQKVEFKASP 180  
 QY 181 FLVGDDCVSIMNGTDEGSAWITINFLTGLSKTPGGSSVGMLDLGGGSTQIAFLPRVGT 240  
 DB 181 FLVGDDCVSIMNGTDEGSAWITINFLTGLSKTPGGSSVGMLDLGGGSTQIAFLPRVGT 240  
 QY 241 LQASPPGYLTALRMFRTYKLSYSLGLGLMSARLAILGGVEGPAKDKELVSPCLSP 300  
 DB 241 LQASPPGYLTALRMFRTYKLSYSLGLGLMSARLAILGGVEGPAKDKELVSPCLSP 300  
 QY 301 SFGGEWEAEVTVYRVSGQAAASLHELCAARVSEVLQNRVHRTVEEVKHYDFVAFSYDDL 360  
 DB 301 SFGGEWEAEVTVYRVSGQAAASLHELCAARVSEVLQNRVHRTVEEVKHYDFVAFSYDDL 360  
 QY 361 AAGVGLIDAKGSLVVGDFEIAAKVCTLETPOSPSPFSCMDLTYSVLLLOEGFPFRS 420  
 DB 361 AAGVGLIDAKGSLVVGDFEIAAKVCTLETPOSPSPFSCMDLTYSVLLLOEGFPFRS 420  
 QY 421 KVLKTRKIDNVETSWALGAIFHYIDSLNRQKSPAS 456  
 DB 421 KVLKTRKIDNVETSWALGAIFHYIDSLNRQKSPAS 456

RESULT 2  
 AAB72241  
 ID AAB72241 standard; protein; 484 AA.  
 XX  
 AC AAB72241;  
 XX

DT 14-MAY-2001 (first entry)  
 XX Human CD39 like protein CD39-L2 amino acid sequence.  
 DE Human CD39 like protein CD39-L2 amino acid sequence.  
 KW Human CD39-like protein; apyrase; NDPase; platelet function inhibitor;  
 KW myocardial infarction; cerebral ischaemia; angina; arterial thrombosis;  
 KW cerebral artery thrombosis; platelet aggregation; inflammation;  
 KW apoptosis; autoimmune disorder; neurological disorder;  
 KW Alzheimer's disease; Parkinson's disease; cancer; CD39-L2.  
 XX Homo sapiens.  
 OS WO200110205-A1.  
 PN 15-FEB-2001.  
 PD 09-AUG-2000; 2000WO-US021790.  
 PF 09-AUG-1999; 99US-00370265.  
 PR 11-JAN-2000; 2000US-00481238.  
 PR 25-APR-2000; 2000US-00557800.  
 PR 26-MAY-2000; 2000US-00583231.  
 PR 30-JUN-2000; 2000US-00608285.  
 XX (HYSE-) HYSEQ INC.  
 PA Ford J, Mulero JJ, Yeung G;  
 FI WPI; 2001-147489/15.  
 XX N-PSDB; AAF63386.  
 DR Polynucleotides encoding human CD39-like polypeptides, with apyrase  
 XX and/or NDPase activity, which are useful in the treatment of pathological  
 PT conditions caused by thrombosis (e.g. myocardial infarction) and  
 PT inflammatory disorders.  
 PS Claim 39; Page 162-164; 203pp; English.  
 XX This invention relates to polynucleotides encoding human CD39-like  
 CC polypeptides with apyrase and/or NDPase activity. The polypeptides having  
 CC ATPase, including NDPase, activity are useful for inhibiting platelet  
 CC function and can therefore be used in the prophylaxis or treatment of  
 CC pathological conditions caused by or involving thrombosis or excessive  
 CC coagulation or excessive platelet aggregation, such as myocardial  
 CC infarction, cerebral ischaemia, angina, arterial thrombosis, cerebral  
 CC artery thrombosis or intracardiac thrombosis, and conditions associated  
 CC with venous thrombosis. CD39-L4 and CD39-L2 polypeptides are useful in  
 CC modulating disease states (including platelet aggregation, inflammation  
 CC and apoptosis) associated with ADP or other purinergic signalling by  
 CC reducing the levels of NDPs. The polypeptides are also useful for  
 CC prophylaxis or treatment of inflammation related disorders, such as  
 CC disorders involving sepsis or systemic inflammatory response syndrome or  
 CC SIRS (and associated conditions such as fever, tachycardia, tachypnea,  
 CC cytokine overstimulation); autoimmune disorders such as thrombosis,  
 CC atherosclerosis, acute pancreatitis, dermatitis, including psoriasis,  
 CC cirrhosis, reperfusion injury, asthma, multiple sclerosis, arthritis;  
 CC neurological disorders including neurodegenerative diseases, epilepsy,  
 CC depression, Alzheimer's disease, Parkinson's disease, Huntington's  
 CC disease, and amyotrophic lateral sclerosis; and cancer. The present  
 CC sequence represents human CD39 like protein CD39-L2  
 XX Sequence 484 AA;

Query Match 100.0%; Score 2364; DB 4; Length 484;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-216;  
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKISNHGSLRVAKVAYPLGLCVGFIYVAYIKWHRATATQAFSTRAAPGAWGQQA 60  
 DB 29 MRKISNHGSLRVAKVAYPLGLCVGFIYVAYIKWHRATATQAFSTRAAPGAWGQQA 88  
 QY 61 SPLGTADGHEVYIGMFDAGSGTGRVHVQFTRPPRPTTLTHETFKAVKPGLSAYADD 120

Db 89 SPLGTADGHEVFGYIMFDAGSTGTRVHVFQFTRPREPTLTTHETFKAVKPGLSAYADD 148  
QY 121 VSKSAQGIREFLDVAKODIPDPFWKATPLVLKATAGRLPGKAKQLQNKVEFKASP 180  
Db 149 VSKSAQGIREFLDVAKODIPDPFWKATPLVLKATAGRLPGKAKQLQNKVEFKASP 208  
QY 181 FLVGGDCVSIINMGDEBVSAMITINFLTGSLLKTPGSSVGMLDLGGSTQIAPLPRVEGT 240  
Db 209 FLVGGDCVSIINMGDEBVSAMITINFLTGSLLKTPGSSVGMLDLGGSTQIAPLPRVEGT 268  
QY 241 LQASPPGYTLALRMFNRTYKLSYSYGLGLMSARLALIGVSGQPAKGEKELVSPLSP 300  
Db 269 LQASPPGYTLALRMFNRTYKLSYSYGLGLMSARLALIGVSGQPAKGEKELVSPLSP 328  
QY 301 SFKGEWEHAETVTVSGOKAASLHELCAARVSEVLQNRVHRTBEVKHVDFFAFSYTYDL 360  
Db 329 SFKGEWEHAETVTVSGOKAASLHELCAARVSEVLQNRVHRTBEVKHVDFFAFSYTYDL 388  
QY 361 AAGVGLIDAEKGGSLVVGDEFEIAKVCRTLETQPOSSPSCMDLTVYSLLQEFKPPRS 420  
Db 389 AAGVGLIDAEKGGSLVVGDEFEIAKVCRTLETQPOSSPSCMDLTVYSLLQEFKPPRS 448  
QY 421 KYLKLTRKIDNVTSMALGALFHYIDSLNRKSPAS 456  
Db 449 KYLKLTRKIDNVTSMALGALFHYIDSLNRKSPAS 484

RESULT 3  
AB06124  
ID ABB06124 standard; protein; 463 AA.  
AC ABB06124;  
XX 10-MAY-2002 (first entry)  
DT Human NS protein sequence SEQ ID NO:216.  
DE  
XX Human, cytosol; osteopathic; gynaecological; neuroprotective;  
KM antitubercular; antitubercular; antiparasitic; ophthalmological; anti-HIV;  
KM vasorelaxant; antitubercular; antitubercular; dermatological;  
KM anorectic; muscular; antitubercular; cardiovascular; antitubercular;  
KM antifibrinolytic; hypotensive; antitubercular; immunomodulator; cardiac;  
KM anticonvulsant; antidiabetic; tranquilizer; antidepressant; neuroleptic;  
KM gastroenteric; antitubercular; antitubercular; cerebroprotective; nootropic;  
KM contraceptive; vaccine; gene therapy; cancer; osteoporosis; dystonia;  
KM endometriosis; degenerative disease; multiple sclerosis; psoriasis;  
KM rheumatoid arthritis; cataract; restenosis; atherosclerosis; glaucoma;  
KM inflammation; skin disorder; obesity; muscular dystrophy; AIDS;  
KM infertility; cardiovascular disease; coagulation disease; hypertension;  
KM ischaemia; asthma; immune disease; epilepsy; angina; neurodegeneration;  
KM diabetes; anxiety; depression; schizophrenia; viral disease; stroke;  
KM gastric ulcer; Alzheimer's disease.  
XX Homo sapiens.  
OS  
XX WO200206315-A2.  
FN  
XX 24-JAN-2002.  
PD  
XX 17-JUL-2001; 2001WO-IL000653.  
PF  
XX 18-JUL-2000; 2000IL-00137345.  
PR  
XX 15-DEC-2000; 2000IL-00140354.  
XX  
XX (COMP-) COMPUGEN LTD.  
PA  
XX Mintz L, Freilich S, Bernstein J;  
PI  
XX WPI; 2002-155037/20.  
DR  
XX N-PSDB; ABL39778.  
XX  
XX One hundred and twenty eight novel nucleic acid sequences, useful for  
PT treating and diagnosing e.g. cancer, asthma and Alzheimer's.

XX  
PS Claim 6, Page 251-253; 290pp; English.  
XX  
XX ABL39691 to ABL39818 represent novel human nucleic acid sequences  
CC encoding the proteins given in ABB06124 to ABB06144. The novel sequences  
CC (NS) can have cytosol, osteopathic, gynaecological, neuroprotective,  
CC antitubercular, antitubercular, antiparasitic, ophthalmological, vaccine,  
CC vasorelaxant, antitubercular, antitubercular, dermatological,  
CC anorectic, muscular, antitubercular, hypotensive, cardiovascular, cardiac,  
CC anticonvulsant, antifibrinolytic, hypotensive, antitubercular, cardiac,  
CC immunomodulator, anticonvulsant, antidiabetic, tranquilizer, antitubercular,  
CC antidepressant, gastroenteric, antitubercular, cerebroprotective,  
CC nootropic and contraceptive activities. The NS can be used in vaccines,  
CC gene therapy and antisense therapy. Nucleic acids, expression vectors and  
CC antibodies from the present invention can be used for treating and  
CC diagnosing e.g. cancer, osteoporosis, endometriosis, degenerative  
CC diseases, dystonia, multiple sclerosis, rheumatoid arthritis, psoriasis,  
CC cataracts, restenosis, atherosclerosis, inflammation, skin disorders,  
CC glaucoma, obesity, muscular dystrophy, AIDS, infertility, cardiovascular  
CC disease, coagulation disease, ischaemia, hypertension, asthma, immune  
CC disease, epilepsy, angina, neurodegeneration, diabetes, anxiety,  
CC depression, schizophrenia, viral disease, gastric ulcers, stroke,  
CC Alzheimer's disease and as a contraceptive  
XX  
SQ Sequence 463 AA:  
Query Match 99.9%; Score 2361; DB 5; Length 463;  
Best Local Similarity 99.8%; Pred. No. 1.1e-215;  
Matches 455; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MKKISHGSLRVAKAYPLGLCVGFIVVAYIKMRAATATQAFSITRAPCARMGQAAH 60  
Db 8 MKKISHGSLRVAKAYPLGLCVGFIVVAYIKMRAATATQAFSITRAPCARMGQAAH 67  
QY 61 SPLGTADGHEVFGYIMFDAGSTGTRVHVFQFTRPREPTLTTHETFKAVKPGLSAYADD 120  
Db 68 SPLGTADGHEVFGYIMFDAGSTGTRVHVFQFTRPREPTLTTHETFKAVKPGLSAYADD 127  
QY 121 VSKSAQGIREFLDVAKODIPDPFWKATPLVLKATAGRLPGKAKQLQNKVEFKASP 180  
Db 128 VSKSAQGIREFLDVAKODIPDPFWKATPLVLKATAGRLPGKAKQLQNKVEFKASP 187  
QY 181 FLVGGDCVSIINMGDEBVSAMITINFLTGSLLKTPGSSVGMLDLGGSTQIAPLPRVEGT 240  
Db 188 FLVGGDCVSIINMGDEBVSAMITINFLTGSLLKTPGSSVGMLDLGGSTQIAPLPRVEGT 247  
QY 241 LQASPPGYTLALRMFNRTYKLSYSYGLGLMSARLALIGVSGQPAKGEKELVSPLSP 300  
Db 248 LQASPPGYTLALRMFNRTYKLSYSYGLGLMSARLALIGVSGQPAKGEKELVSPLSP 307  
QY 301 SFKGEWEHAETVTVSGOKAASLHELCAARVSEVLQNRVHRTBEVKHVDFFAFSYTYDL 360  
Db 308 SFKGEWEHAETVTVSGOKAASLHELCAARVSEVLQNRVHRTBEVKHVDFFAFSYTYDL 367  
QY 361 AAGVGLIDAEKGGSLVVGDEFEIAKVCRTLETQPOSSPSCMDLTVYSLLQEFKPPRS 420  
Db 368 AAGVGLIDAEKGGSLVVGDEFEIAKVCRTLETQPOSSPSCMDLTVYSLLQEFKPPRS 427  
QY 421 KYLKLTRKIDNVTSMALGALFHYIDSLNRKSPAS 456  
Db 428 KYLKLTRKIDNVTSMALGALFHYIDSLNRKSPAS 463

RESULT 4  
ABJ04657  
ID ABJ04657 standard; protein; 467 AA.  
AC ABJ04657;  
XX 11-OCT-2002 (first entry)  
DT  
XX Protein of NOVX 15a SEQ ID No 36.  
DE  
XX

KW Cytostatic; antidiabetic; anorectic; metabolic; nootropic; antilipemic;  
 KW neuroprotective; antiparkinsonian; anticonvulsant; cerebroprotective;  
 KW tranquilliser; neuroleptic; antidiabetic; antiulcer; antiinflammatory;  
 KW anti-HIV; antiallergic; antirheumatic; antiarthritic; NOVX; diabetes;  
 KW metabolic disorder; obesity, infectious disease; Alzheimer's disease;  
 KW anorexia; neurodegenerative disorder; Parkinson's disorder; obesity;  
 KW immune disorder; haematopoietic disorder; dyslipidaemia; chronic disease;  
 KW metabolic syndrome X; wasting disorder; cancer; neurological disorder;  
 KW epilepsy; stroke; mental disorder; schizophrenic disorders; goiter;  
 KW vesicular transport; cystic fibrosis; gastrointestinal disorder;  
 KW diabetes mellitus; ulcerative colitis; AIDS; allergic reaction;  
 KW multiple sclerosis; rheumatoid arthritis; transgenic animal;  
 KW gene therapy.

XX Unidentified.  
 XX WO200246409-A2.  
 XX 13-JUN-2002.  
 XX 06-DEC-2001; 2001WO-US046586.

XX 06-DEC-2000; 2000US-0251660P.  
 PR 12-DEC-2000; 2000US-0255029P.  
 PR 08-JAN-2001; 2001US-0260326P.  
 PR 24-JAN-2001; 2001US-0263800P.  
 PR 20-FEB-2001; 2001US-0269942P.  
 PR 24-APR-2001; 2001US-0286183P.  
 PR 20-AUG-2001; 2001US-0313627P.  
 PR 12-SEP-2001; 2001US-0318712P.

XX (CURA-) CURAGEN CORP.

XX Guo X, Li L, Patturajan M, Shimkets RA, Casman SJ, Malyankar UM;  
 PI Tchernev VT, Vernet CAM, Spytek KA, Shenoy SG, Alsbrook JP;  
 PI Edinger S, Peyman JA, Stone DJ, Ellerman K, Gangolli EA, Boldog FI;  
 PI Colman SD, Eisen AJ, Liu X, Padigaru M, Spaderna SK, Zerhusen BD;

XX WPI: 2002-547774/58.  
 DR N-PSDB; ABT05470.

XX Novel isolated polypeptide, designated NOVX, useful for treating or  
 PT preventing cancer, diabetes, obesity, dyslipidemia, anorexia, and  
 PT metabolic, neurodegenerative, immune and hematopoietic disorders.

XX Claim 1; Page 140; 421pp; English.

XX The invention relates to an isolated polypeptide, designated NOVX,  
 CC comprising a sequence fully defined in the specification. The isolated  
 CC protein, its encoding polynucleotide or an antibody created from the  
 CC protein is useful in the manufacture of a medicament for treating a  
 CC syndrome associated with a human disease, preferably a NOVX-associated  
 CC disorder, or for treating or preventing a NOVX-associated disorder in a  
 CC subject, preferably human. The isolated protein, its encoding  
 CC polynucleotide or an antibody created from the protein are also useful  
 CC for treating or preventing metabolic disorders, diabetes, obesity,  
 CC infectious disease, anorexia, neurodegenerative disorder, Alzheimer's  
 CC disease, Parkinson's disorder, immune disorders, haematopoietic  
 CC disorders, and various dyslipidaemias, metabolic disturbances associated  
 CC with obesity, the metabolic syndrome X, wasting disorders associated with  
 CC chronic diseases, and cancer. The isolated protein, its encoding  
 CC polynucleotide or an antibody created from the protein are useful for  
 CC treating or preventing neurological disorders such as epilepsy, stroke,  
 CC mental disorders including mood, anxiety, schizophrenic disorders,  
 CC disorders of vesicular transport such as cystic fibrosis, diabetes  
 CC mellitus, goiter, gastrointestinal disorders including ulcerative  
 CC colitis, other conditions associated with abnormal vesicle trafficking  
 CC including AIDS, allergic reactions, multiple sclerosis and rheumatoid  
 CC arthritis. A cell comprising the vector of the invention is useful for  
 CC producing non-human transgenic animals. The polynucleotide of the  
 CC invention can be used to treat disorders by gene therapy. This sequence  
 CC represents one of the isolated NOVX proteins of the invention

SQ Sequence 467 AA;  
 Query Match 99.9%; Score 2361; DB 5; Length 467;  
 Best Local Similarity 99.8%; Pred. No. 1.1e-215;  
 Matches 455; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MRKISNHGSLRVAKVAYPLGLCVGVYIYVAYIKWHRATATQAFPSITRAAPGARWQQQAH 60  
 DB 12 MRKISNHGSLRVAKVAYPLGLCVGVYIYVAYIKWHRATATQAFPSITRAAPGARWQQQAH 71  
 QY 61 SPLGTAADGHEVFGIMFDAGSGTGVHVQFTRPREPTLTHTETPKAVKPCLSAYADD 120  
 DB 72 SPLGTAADGHEVFGIMFDAGSGTGVHVQFTRPREPTLTHTETPKAVKPCLSAYADD 131  
 QY 121 VEKSAQGIREDLDVAKODIPDFWKATPLVLKATAGLRLLPGKAKQKLLQKVEVFKASP 180  
 DB 132 VEKSAQGIREDLDVAKODIPDFWKATPLVLKATAGLRLLPGKAKQKLLQKVEVFKASP 191  
 QY 181 FLYGDDCVSIMGTDGVSAMITINELTSLKTPGGSSVGMLDLGGSGTOIAFLPRVEGT 240  
 DB 192 FLYGDDCVSIMGTDGVSAMITINELTSLKTPGGSSVGMLDLGGSGTOIAFLPRVEGT 251  
 QY 241 LQASPPGYLTALRMENRTYKLYSYSLGLGMSARLAILGGVGGQPAKDGKELVSPCLSP 300  
 DB 252 LQASPPGYLTALRMENRTYKLYSYSLGLGMSARLAILGGVGGQPAKDGKELVSPCLSP 311  
 QY 301 SFKGWEHAETVTVRSQKAAASLHELCAARSEVLQNRVHRTTEVKHVDYFAFSYYDL 360  
 DB 312 SFKGWEHAETVTVRSQKAAASLHELCAARSEVLQNRVHRTTEVKHVDYFAFSYYDL 371  
 QY 361 AAGVGLLDAEKGSLVVGDFEIAAKVYCRLETQPOSSPFCMDLTYSVLLQLQEFPPRS 420  
 DB 372 AAGVGLLDAEKGSLVVGDFEIAAKVYCRLETQPOSSPFCMDLTYSVLLQLQEFPPRS 431  
 QY 421 KVLKLTPIKIDNVETSWALGAIFHYIISLNKQKSPAS 456  
 DB 432 KVLKLTPIKIDNVETSWALGAIFHYIISLNKQKSPAS 467  
 RESULT 5  
 HAM93929  
 ID AAM93929 standard; protein; 456 AA.  
 XX AC AAM93929;  
 XX DT 06-NOV-2001 (first entry)  
 XX DE Human polypeptide, SEQ ID NO: 4100.  
 XX KW Human; full length cDNA; cDNA synthesis; oligo-capping.  
 XX OS Homo sapiens.  
 XX FN EP1130094-A2.  
 XX PD 05-SEP-2001.  
 XX PF 07-JUL-2000; 2000EP-00114089.  
 XX PR 08-JUL-1999; 99JP-00194486.  
 XX PR 11-JAN-2000; 2000JP-00118774.  
 XX PR 02-MAY-2000; 2000JP-00183765.  
 XX PA (HELI-) HELIX RES INST.  
 XX PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;  
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
 XX WPI: 2001-524255/58.  
 XX DR N-PSDB; AAK94892.  
 XX PT 830 Primers useful for synthesizing full length cDNA clones and their use  
 PT in genetic manipulation.

XX Claim 8; SEQ ID NO 4100; 1380bp + Sequence Listing; English.  
PS  
XX The invention relates to primers for synthesizing full length cDNA  
CC clones. 830 cDNA molecules encoding a human protein have been isolated  
CC and nucleotide sequences of 5' - and 3' - ends of the cDNA molecules have  
CC been determined. Primers for synthesizing the full length cDNA are useful  
CC for clarifying the function of the protein encoded by the cDNA. The full  
CC length clones were obtained by construction of full length enriched cDNA  
CC libraries that were synthesised by the oligo-capping method. The primers  
CC enable the production of the full length cDNA easily without any special  
CC methods. The present sequence is a polypeptide encoded by a full length  
CC human cDNA of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in CD-ROM  
CC format directly from EPO  
XX  
SQ Sequence 456 AA;  
Query Match 99.5%; Score 2351; DB 4; Length 456;  
Best Local Similarity 99.3%; Pred. No. 9.3e-215;  
Matches 453; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MRKISNHSGLRVAKAYAPLGLCVGFIVAYIKWHRATATQAFSITRAAPGARWGQAH 60  
DB 1 MRKISNHSGLRVAKAYAPLGLCVGFIVAYIKWHRATATQAFSITRAAPGARWGQAH 60  
QY 61 SPLGTADGHEVFYIGIMFDAGSTGRVHVFQTRPPRETPTLTHETFRKAVPGISAYAD 120  
DB 61 SPLGTADGHEVFYIGIMFDAGSTGRVHVFQTRPPRETPTLTHETFRKAVPGISAYAD 120  
QY 121 VEKSAQGRRELDVAKODIPDFMKATPLVATAGARILPGEKAOGLLOKVEVFAASP 180  
DB 121 VEKSAQGRRELDVAKODIPDFMKATPLVATAGARILPGEKAOGLLOKVEVFAASP 180  
QY 121 VEKSAQGRRELDVAKODIPDFMKATPLVATAGARILPGEKAOGLLOKVEVFAASP 180  
DB 181 FLVGGDDCVSINMGTDGVSAMITINFLTSLKTPGSSVGMGLDGGSTQIAFLPRVEGT 240  
DB 181 FLVGGDDCVSINMGTDGVSAMITINFLTSLKTPGSSVGMGLDGGSTQIAFLPRVEGT 240  
QY 241 LQASPGVLTALRMFNRTYKYSYSLGLGMSARLILIGVEGQPKDGEVSPCLSP 300  
DB 241 LQASPGVLTALRMFNRTYKYSYSLGLGMSARLILIGVEGQPKDGEVSPCLSP 300  
QY 301 SFKGEWEHAETVYVSGQKAAASHLCAARVSEVLQNRVHRTVEVHVDVFAYFSYYDL 360  
DB 301 SFKGEWEHAETVYVSGQKAAASHLCAARVSEVLQNRVHRTVEVHVDVFAYFSYYDL 360  
QY 361 AAGVGLIDAEKGGSLVGDPEFIANKYVCTTLETQPOSSPFCMDLTYVSLILOEFGPPRS 420  
DB 361 AAGVGLIDAEKGGSLVGDPEFIANKYVCTTLETQPOSSPFCMDLTYVSLILOEFGPPRS 420  
QY 421 KYLKLTRKIDIVETSMALGFHYIDSINROKSPAS 456  
DB 421 KYLKLTRKIDIVETSMALGFHYIDSINROKSPAS 456  
RESULT 6  
ADCL4220  
ID ADCL4220 standard; protein; 450 AA.  
XX  
XX ADCL4220;  
XX  
XX 18-DEC-2003 (first entry)  
XX  
XX Human enzyme ENZM-26.  
XX  
XX enzyme; human; ENZM; cytosolic; antiarteriosclerotic; antidiabetic;  
XX anticonvulsant; nootropic; neuroprotective; cerebroprotective; anti-HIV;  
XX antiatherogenic; antiinflammatory; thymimetic; gene therapy;  
XX cell proliferative disorder; endocrine disorder; neurological disorder;  
XX immune system disorder; inflammatory disorder; developmental disorder;  
XX reproductive disorder; vesicle-trafficking disorder; infection.  
OS Homo sapiens.

XX  
PN WO2003042357-A2.  
XX  
XX 22-MAY-2003.  
XX  
XX 26-SEP-2002; 2002WO-US011096.  
XX  
XX 26-SEP-2001; 2001US-0326388P.  
XX  
XX 12-OCT-2001; 2001US-0328979P.  
XX  
XX 19-OCT-2001; 2001US-0346034P.  
XX  
XX 26-OCT-2001; 2001US-0348284P.  
XX  
XX 08-NOV-2001; 2001US-0338048P.  
XX  
XX 16-NOV-2001; 2001US-0332340P.  
XX  
XX 14-DEC-2001; 2001US-0340357P.  
XX  
XX 29-MAR-2002; 2002US-0368722P.  
XX  
XX 29-MAR-2002; 2002US-0368799P.  
XX  
XX 17-MAY-2002; 2002US-0381558P.  
XX  
XX 07-JUN-2002; 2002US-0387119P.  
XX  
XX 21-JUN-2002; 2002US-0390662P.  
XX  
XX (INCYTE GENOMICS INC.  
XX  
XX Yang J, Lu DM, Yue H, Elliott VS, Warren BA, Duggan BM;  
XX Forsythe JD, Lee EA, Hafalia AUA, Ramkumar J, Chawla NK, Baughn MR;  
XX Becha SD, Gorvad AE, Tran UK, Li JX, Yao MG, Ison CH, Griffin JA;  
XX Lee SY, Chang H, Emerling BM, Tang YT, Lal PG, Kable AE, A;  
XX Margulis JP, Jiang X, Jackson AA, Zebadjadian Y, Sarrnakar A,  
XX Wilson AD, Jin P, Richardson TW, Bhatia U, Burrill JD, Lee S;  
XX Blake JJ, Ho A, Zheng W, Gao J;  
XX  
XX WPI; 2003-449567/42.  
XX  
XX N-P-SDB; ADCL4273.  
XX  
XX New human enzymes (ENZM), useful for diagnosing, treating and preventing  
XX diseases or conditions associated with the aberrant ENZM expression e.g.  
XX cancer, diabetes, epilepsy, or infections.  
XX  
XX  
PS Claim 1; SEQ ID NO 26; 416pp; English.  
XX  
XX The invention relates to a novel isolated human enzyme (ENZM)  
XX polypeptide. A polypeptide of the invention has cytosolic,  
XX antiarteriosclerotic, antidiabetic, anticonvulsant, nootropic,  
XX neuroprotective, cerebroprotective, anti-HIV, antiatherogenic,  
XX antiinflammatory, and thymimetic activity. A polynucleotide encoding a  
XX polypeptide of the invention may have a use in gene therapy. The  
XX polypeptide and polynucleotides are useful in diagnosing, treating and  
XX preventing diseases or conditions associated with the decreased  
XX expression or overexpression of ENZM, such as cell proliferative (e.g.  
XX cancer, atherosclerosis), endocrine (e.g. diabetes), neurological (e.g.  
XX epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,  
XX allergies), developmental (e.g. Hypothyroidism, Cushing's syndrome),  
XX reproductive and vesicle-trafficking disorders, or infections. These are  
XX also useful in assessing the effects of exogenous compounds on the  
XX expression of nucleic acid and amino acid sequences of ENZM. The ENZM or  
XX its fragments are useful in screening compounds for effectiveness as  
XX agonist or antagonist of the polypeptides, or in altering the expression  
XX of the target polynucleotide and compounds that specifically bind to or  
XX modulate the activity of the polypeptide. The microarray is useful in  
XX monitoring or measuring protein-protein interactions, drug-target  
XX interactions, and gene expression profiles. The sequences shown in  
XX ADCL4195-ADCL4247 represent ENZM proteins of the invention.  
XX  
XX  
SQ Sequence 450 AA;  
Query Match 91.0%; Score 2152; DB 7; Length 450;  
Best Local Similarity 92.3%; Pred. No. 8.3e-196;  
Matches 421; Conservative 1; Mismatches 0; Indels 34; Gaps 1;  
QY 1 MRKISNHSGLRVAKAYAPLGLCVGFIVAYIKWHRATATQAFSITRAAPGARWGQAH 60  
DB 29 MRKISNHSGLRVAKAYAPLGLCVGFIVAYIKWHRATATQAFSITRAAPGARWGQAH 88  
QY 61 SPLGTADGHEVFYIGIMFDAGSTGRVHVFQTRPPRETPTLTHETFRKAVPGISAYAD 120



XX MO200246409-A2.  
 XX 13-JUN-2002.  
 XX 06-DEC-2001; 2001WO-US046586.  
 XX 06-DEC-2000; 2000US-0251660P.  
 XX 12-DEC-2000; 2000US-0255029P.  
 XX 08-JAN-2001; 2001US-0260326P.  
 XX 24-JAN-2001; 2001US-0263800P.  
 XX 20-FEB-2001; 2001US-0269942P.  
 XX 24-APR-2001; 2001US-0286183P.  
 XX 20-AUG-2001; 2001US-0313627P.  
 XX 12-SEP-2001; 2001US-0318712P.  
 XX (CURA-) CURAGEN CORP.  
 XX Guo X, Li L, Paturajan M, Shinkets RA, Casman SJ, Malyankar UM,  
 XX Tchernov ST, Vernet CAM, Szytek KA, Shenoy SG, Alsbrock JP,  
 XX Edinger S, Peyman JA, Stone DJ, Ellerman K, Gangoli EA, Boldog PL,  
 XX Colman SD, Eisen AJ, Liu X, Padigaru M, Spaderna SK, Zernhusen BD;  
 XX MPI: 2002-547774/58.  
 XX N-PSDB; ABT05471.  
 XX Novel isolated polypeptide, designated NOVX, useful for treating or  
 XX preventing cancer, diabetes, obesity, dyslipidemia, anorexia, and  
 XX metabolic, neurodegenerative, immune and hematopoietic disorders.  
 XX Claim 1; Page 141; 421pp; English.  
 XX The invention relates to an isolated polypeptide, designated NOVX,  
 XX comprising a sequence fully defined in the specification. The isolated  
 XX protein, its encoding polynucleotide or an antibody created from the  
 XX protein is useful in the manufacture of a medicament for treating a  
 XX syndrome associated with a human disease, preferably a NOVX-associated  
 XX disorder, or for treating or preventing a NOVX-associated disorder in a  
 XX subject, preferably human. The isolated protein, its encoding  
 XX polynucleotide or an antibody created from the protein are also useful  
 XX for treating or preventing metabolic disorders, diabetes, obesity,  
 XX infectious disease, anorexia, neurodegenerative disorder, Alzheimer's  
 XX disease, Parkinson's disorder, immune disorders, haematopoietic  
 XX disorders, and various dyslipidemias, metabolic disturbances associated  
 XX with obesity, the metabolic syndrome X, wasting disorders associated with  
 XX chronic diseases, and cancer. The isolated protein, its encoding  
 XX polynucleotide or an antibody created from the protein are useful for  
 XX treating or preventing neurological disorders such as epilepsy, stroke,  
 XX mental disorders including mood, anxiety, schizophrenic disorders,  
 XX disorders of vesicular transport such as cystic fibrosis, diabetes  
 XX mellitus, goiter, gastrointestinal disorders including ulcerative  
 XX colitis, other conditions associated with abnormal vesicle trafficking  
 XX including AIDS, allergic reactions, multiple sclerosis and rheumatoid  
 XX arthritis. A cell comprising the vector of the invention is useful for  
 XX producing non-human transgenic animals. The polynucleotide of the  
 XX invention can be used to treat disorders by gene therapy. This sequence  
 XX represents one of the isolated NOVX proteins of the invention  
 XX Sequence 446 AA;  
 SO Query Match 89.5%; Score 2116; DB 5; Length 446;  
 Best Local Similarity 91.0%; Pred. No. 2,2e-192;  
 Matches 415; Conservative 1; Mismatches 2; Indels 38; Gaps 1;  
 QY 1 MRKISNHSGLRVAVAYPLIGCVFIYVAYIKMHRATATQAFPSITRAAPGARMGQQA 60  
 DB 29 MRKISNHSGLRVAV-----ARMGQQA 50  
 QY 61 SPICGTADGHEVYGVYIMFDAGSTGRVHVFOFTPRPPTPTLTHETFKAVKPGISAVADD 120  
 DB 51 SPLGTADGHEVYGVYIMFDAGSTGRVHVFOFTPRPPTPTLTHETFKALKPGISAVADD 110  
 QY 121 VERAQGIREFLDVAKODIPDFWKATPVLKATAGRLRPGKAQKLLQKVEVFAPAS 180

DB 111 VERAQGIREFLDVAKODIPDFWKATPVLKATAGRLRPGKAQKLLQKVEVFAPAS 170  
 QY 181 FLVGDCCVSIIMNGDEGVSAWITINPLTSLKTPGSSVGMGLGGSTPQIAFLPVEGT 240  
 DB 171 FLVGDCCVSIIMNGDEGVSAWITINPLTSLKTPGSSVGMGLGGSTPQIAFLPVEGT 230  
 QY 241 LQASPPGTLALRMFNRTYKLYSYSLGLGMSARLAILGVEGQPAKDKELVSPCLSP 300  
 DB 231 LQASPPGTLALRMFNRTYKLYSYSLGLGMSARLAILGVEGQPAKDKELVSPCLSP 290  
 QY 301 SFKGEWEHAETTVVSQKAAASHELCARVSEVYLQNRHREVEVQVDFYAFSTYYDL 360  
 DB 291 SFKGEWEHAETTVVSQKAAASHELCARVSEVYLQNRHREVEVQVDFYAFSTYYDL 350  
 QY 361 AAGVGLDAEKGSGLVYGDPEIAKVCRTLETQPSPPSCMDLTYSLLLOEFGPPRS 420  
 DB 351 AAGVGLDAEKGSGLVYGDPEIAKVCRTLETQPSPPSCMDLTYSLLLOEFGPPRS 410  
 QY 421 KVLKLTREKIDNVEFSMALGALFHYIDSLNRQKSPAS 456  
 DB 411 KVLKLTREKIDNVEFSMALGALFHYIDSLNRQKSPAS 446  
 RESULT 9  
 AAB72242  
 ID AAB72242 standard; protein; 471 AA.  
 AC AAB72242;  
 XX 14-MAY-2001 (first entry)  
 DT 14-MAY-2001 (first entry)  
 XX Mature human CD39 like protein CD39-L2 amino acid sequence.  
 DE Human CD39-like protein; apyrase; NDPase; platelet function inhibitor;  
 XX myocardial infarction; cerebral ischemia; angina; arterial thrombosis;  
 KW cerebral artery thrombosis; platelet aggregation; inflammation;  
 KW apoptosis; autoimmune disorder; neurological disorder;  
 KM Alzheimer's disease; Parkinson's disease; cancer; CD39-L2.  
 XX Homo sapiens.  
 OS WO200110205-A1.  
 XX 15-FEB-2001.  
 PD 09-AUG-2000; 2000WO-US021790.  
 XX 09-AUG-1999; 99US-00370265.  
 PR 11-JAN-2000; 2000US-00481238.  
 PR 25-APR-2000; 2000US-00557800.  
 PR 26-MAY-2000; 2000US-00583231.  
 PR 30-JUN-2000; 2000US-00608285.  
 XX (HYSE-) HYSEQ INC.  
 PA Ford J, Mulero JJ, Yeung G;  
 XX MPI: 2001-147489/15.  
 DR N-PSDB; AAF63387.  
 XX Polynucleotides encoding human CD39-like polypeptides, with apyrase  
 PT and/or NDPase activity, which are useful in the treatment of pathological  
 PT conditions caused by thrombosis (e.g. myocardial infarction) and  
 XX inflammatory disorders.  
 XX Claim 53; Page 192-194; 203pp; English.  
 XX This invention relates to polynucleotides encoding human CD39-like  
 XX polypeptides with apyrase and/or NDPase activity. The polypeptides having  
 CC ATPase, including NDPase, activity are useful for inhibiting platelet  
 CC function and can therefore be used in the prophylaxis or treatment of  
 CC pathological conditions caused by or involving thrombosis or excessive

CC coagulation or excessive platelet aggregation, such as myocardial  
 CC infarction, cerebral ischaemia, angina, arterial thrombosis, cerebral  
 CC artery thrombosis or intracardiac thrombosis, and conditions associated  
 CC with venous thrombosis. CD39-L4 and CD39-L2 polypeptides are useful in  
 CC modulating disease states (including platelet aggregation, inflammation  
 CC and apoptosis) associated with ADP or other purinergic signalling by  
 CC reducing the levels of NTPs. The polypeptides are also useful for  
 CC prophylaxis or treatment of inflammation related disorders, such as  
 CC disorders involving sepsis or systemic inflammatory response syndrome or  
 CC SIRS (and associated conditions such as fever, tachycardia, tachypnea,  
 CC cytokine overstimulation); autoimmune disorders such as thrombosis,  
 CC atherosclerosis, acute pancreatitis, dermatitis, including psoriasis,  
 CC cirrhosis, reperfusion injury, asthma, multiple sclerosis, arthritis,  
 CC neurological disorders including neurodegenerative diseases, epilepsy,  
 CC depression, Alzheimer's disease, Parkinson's disease, Huntington's  
 CC disease, and amyotrophic lateral sclerosis; and cancer. The present  
 CC sequence represents mature D39 like protein CD39-L2

XX  
 SQ Sequence 471 AA;

Query Match 84.7%; Score 2003; DB 4; Length 471;  
 Best Local Similarity 98.5%; Pred. No. 1.4e-181;  
 Matches 388; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 MRKISNHGSLRVAKVAYPLGLGVFIYVAYIKWHDATATQAFSSITRAAPGARWQQAH 60  
 DB 29 MRKISNHGSLRVAKVAYPLGLGVFIYVAYIKWHDATATQAFSSITRAAPGARWQQAH 88  
 QY 61 SPLGTAADGHEVFYGMEDAGSTGTRVHVFOFTPRETPTLTHTFFKAVKGLSNAYADD 120  
 DB 89 SPLGTAADGHEVFYGMEDAGSTGTRVHVFOFTPRETPTLTHTFFKAVKGLSNAYADD 148  
 QY 121 VEKSAQGIREDLLVAKQDIPFDKATPLVLKATAGLRLPGEKAKQLKQKVEFKASP 180  
 DB 149 VEKSAQGIREDLLVAKQDIPFDKATPLVLKATAGLRLPGEKAKQLKQKVEFKASP 208  
 QY 181 FLVGGDDCVSINMGTDGVSAMITINFLTGLTPGGSSVGMGLDGGGSTQIAFLPRVEGT 240  
 DB 209 FLVGGDDCVSINMGTDGVSAMITINFLTGLTPGGSSVGMGLDGGGSTQIAFLPRVEGT 268  
 QY 241 LQASPPGYLTALRMFNRKYKLYSYVLGLGLMSARLAILGGVEGQPAKDKGLVSPCLSP 300  
 DB 269 LQASPPGYLTALRMFNRKYKLYSYVLGLGLMSARLAILGGVEGQPAKDKGLVSPCLSP 328  
 QY 301 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTVEVGVDFVAFSYDDL 360  
 DB 329 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTVEVGVDFVAFSYDDL 388  
 QY 361 AAGVGLIDAEKGGSLVVGDFEIAAKVVCRTLEQ 394  
 DB 389 AAGVGLIDAEKGGSLVVGDFEIAAKVGGSHLERE 422

RESULT 10

AA444849  
 ID AA444849 standard; protein; 428 AA.

AC AA444849;

XX 18-MAY-2000 (first entry)

XX Human CD39-L4 protein.

XX CD39-L4; human; apyrase; nucleotide diphosphatase; NDase;

XX ATP diphosphohydrolase; ATPase; adenosine diphosphate; ADP; treatment;  
 XX platelet aggregation; antithrombotic; thrombosis; myocardial infarction;  
 XX cerebral ischaemia; angina; vascular graft; extracorporeal circulation;  
 XX molecular weight marker; nutritional supplement; tumour; prevention;  
 XX drug targeting; Apyrase Conserved Region; ACR.

OS Homo sapiens.

XX Key Location/Qualifiers

FH

FT Peptide 1. .22  
 FT /label= Leader\_peptide  
 FT Protein 23. .428  
 FT /label= Mature human CD39-L4 protein  
 FT /note= "Homologous to human and murine CD39"  
 FT Binding-site 54. .58  
 FT /label= ATP\_Binding\_region  
 FT Region 129. .134  
 FT /label= Apyrase\_Conserved\_Region  
 FT Region 169. .173  
 FT /label= Apyrase\_Conserved\_Region  
 FT Region 199. .206  
 FT /note= "Conserved motif in ATPDases"  
 PN WO200004041-A2.  
 XX 27-JAN-2000.  
 XX 16-JUL-1999; 99WO-US016180.  
 XX 16-JUL-1998; 98US-00118205.  
 PR 24-JUL-1998; 98US-00122449.  
 PR 04-FEB-1999; 99US-00244444.  
 PR 19-MAR-1999; 99US-00273447.  
 PR 09-JUL-1999; 99US-00350836.  
 XX (HYSE-) HYSEQ INC.  
 XX Ford J, Mulero J;  
 XX WPI; 2000-192397/16.  
 DR N-PSDB; AAZ50356, AAZ50359.  
 XX New nucleic acid encoding human CD39-like protein, useful for treating  
 XX and preventing thrombotic disease.  
 XX Claim 15; Fig 2; 125pp; English.  
 XX The present amino acid sequence is the CD39-L4 protein, an apyrase and/or  
 XX nucleotide diphosphatase (NDase). It is isolated from the human foetal  
 XX liver-spleen cDNA library, b2HFLS20W. It is a soluble ATP  
 XX diphosphohydrolase (ATPDase) and is involved in the hydrolysis of  
 XX adenosine diphosphate (ADP), the agonist that causes platelet  
 XX aggregation. CD39-L4 protein has 30% and 80% homology to human and murine  
 XX CD39. It has platelet aggregation inhibition and antithrombotic activity.  
 XX CD39-L4 is used to treat or prevent thrombosis, myocardial infarction,  
 XX cerebral ischaemia and angina. It is also used in vitro, to maintain  
 XX vascular grafts or during extracorporeal circulation, to hydrolyse NDP,  
 XX as molecular weight markers and as nutritional supplements. It is used to  
 XX identify therapeutic agents that bind and modulate CD39-L4. It is coupled  
 XX to toxins for targeting drugs to tumours or other cells that express CD39  
 XX -L4  
 SQ Sequence 428 AA;  
 Query Match 42.3%; Score 999; DB 3; Length 428;  
 Best Local Similarity 52.4%; Pred. No. 5.5e-86;  
 Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;  
 QY 62 PLGTAADGHEVFYGMEDAGSTGTRVHVFOFTPRETPTLTHTFFKAVKGLSNAYADD 120  
 DB 40 PINVSA--STLYGIMEDAGSTGTRIHVYTFVQMPGQLFLEGEVFDKSVKPGLSAFVDQ 96  
 QY 121 VEKSAQGIREDLLVAKQDIPFDKATPLVLKATAGLRLPGEKAKQLKQKVEFKASP 180  
 DB 97 PKQAEVTVQGLLEVAKQDIPRSHWKTFFVVLKATAGLRLPGEKAKALLFEVKEIFKSP 156  
 QY 181 FLVGGDDCVSINMGTDGVSAMITINFLTGLTPGGSSVGMGLDGGGSTQIAFLPRVEGT 240  
 DB 157 FLVFKGVSIMDGSDEGLAWTVNFLTGLHGRQETVGTLDLGGASTQITFLPQPEKT 216  
 QY 241 LQASPPGYLTALRMFNRKYKLYSYVLGLGLMSARLAILGGVEGQPAKDKGLVSPCLSP 300



Db 217 LEQTPRGYLTFEMENSTYKLYTHSYLGFGLKAAKLATLGLAETE-GTDGHTFRSACLPR 275  
 QY 301 SFKREMEHAETTVYVSGQKAAASHLHCAARVSEVLQNRVHRDEYVNDVFAFSYYDL 360  
 Db 276 WLEAEWIFGVKYQYGGNQBSEVGPPEPCYAEVLNVRVSKLHQPEEVORGSFYAFSYYDR 335  
 QY 361 AAGVGLIDAEKGSILVVGDFEIAKVCRTLETQPSQSPSCMDLTYSLLQOE-FGFPFR 419  
 Db 336 AVDTDMIDYEKGGLTKVEDFERKAREVCNDLNEENTSSPFLCMDLSYITALLDGSFGFAD 395  
 QY 420 SKVTLTKRKIDNVETSMALGALFYHYDSL 448  
 Db 396 STVLTQTKKVNNIETGMALGATFHLQSL 424

RESULT 11  
 AAB72238  
 ID AAB72238 standard; protein; 428 AA.  
 XX  
 AC AAB72238;  
 XX  
 DT 14-MAY-2001 (first entry)  
 XX

Human CD39 like protein CD39-L4 amino acid sequence.

Human CD39-like protein; apyrase; NDase; platelet function inhibitor;  
 myocardial infarction; cerebral ischaemia; angina; arterial thrombosis;  
 cerebral artery thrombosis; platelet aggregation; inflammation;  
 apoptosis; autoimmune disorder; neurological disorder;  
 Alzheimer's disease; Parkinson's disease; cancer; CD39-L4.

Homo sapiens.

MO200110205-A1.

15-FEB-2001.

09-AUG-2000; 2000MO-US021790.

09-AUG-1999; 99US-00370265.

11-JAN-2000; 2000US-00481238.

25-APR-2000; 2000US-00557800.

26-MAY-2000; 2000US-00583231.

30-JUN-2000; 2000US-00608285.

(HYSE-) HYSEQ INC.

Ford J, Mulero JJ, Yeung G;

WPI; 2001-147489/15.

N-PSDB; AAF63383.

Polynucleotides encoding human CD39-like polypeptides, with apyrase

and/or NDase activity, which are useful in the treatment of pathological

conditions caused by thrombosis (e.g. myocardial infarction) and

inflammatory disorders.

Claim 15; Fig 2; 203pp; English.

This invention relates to polynucleotides encoding human CD39-like  
 polypeptides with apyrase and/or NDase activity. The polypeptides having  
 ATPase, including NDase, activity are useful for inhibiting platelet  
 function and can therefore be used in the prophylaxis or treatment of  
 pathological conditions caused by or involving thrombosis or excessive  
 coagulation or excessive platelet aggregation such as myocardial  
 infarction, cerebral ischaemia, angina, arterial thrombosis, cerebral  
 artery thrombosis or intracardiac thrombosis, and conditions associated  
 with venous thrombosis. CD39-L4 and CD39-L2 polypeptides are useful in  
 modulating disease states (including platelet aggregation, inflammation  
 and apoptosis) associated with ADP or other purinergic signalling by  
 reducing the levels of NDase. The polypeptides are also useful for  
 prophylaxis or treatment of inflammation related disorders, such as  
 disorders involving sepsis or systemic inflammatory response syndrome or

CC SIRS (and associated conditions such as fever, tachycardia, tachypnea,  
 CC cytokine overstimulation); autoimmune disorders such as thrombosis,  
 CC atherosclerosis, acute pancreatitis, dermatitis, including psoriasis,  
 CC clithosis, reperfusion injury, asthma, multiple sclerosis, arthritis,  
 CC neurological disorders including neurodegenerative diseases, epilepsy,  
 CC depression, Alzheimer's disease, Parkinson's disease, Huntington's  
 CC disease, and amyotrophic lateral sclerosis; and cancer. The present  
 CC sequence represents human CD39 like protein CD39-L4  
 XX

SO Sequence 428 AA;

Query Match 42.3%; Score 999; DB 4; Length 428;  
 Best Local Similarity 52.4%; Pred. No. 5,5e-86;  
 Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;

QY 62 PLGTADGHEVRYGMPDAGSTGTRVHYFOFT-REPREPTLTHFRAXKRGSLAVYDD 120

Db 40 PLNVSH---STVYGMFDHSGSTGRHIVYTFVQKRPGLPIEGVSPVAFGLSAFVDQ 96

QY 121 VEKSAQGIREFLDVAKODIPDFWKAFTPLVKATAGLRLPGEKAQKLQRYKEVFKASP 180

Db 97 PKGAEFTVQGLELVAKDSIPRSHWKTPLVKATAGLRLPGEKAKALLFEVKEIFRASP 156

QY 181 FLVGDQCVSINMGTEGVSAMTTFNLGSLKTPGSSVGMULDGGSQTQAFPRVSGT 240

Db 157 FLVPRGVSIVMDGSDGELIAVTVNPLTGQHGHRQETVGLDGGASTQTLFQFEKT 216

QY 241 LQASPPGYTLALRNRYTKLYSYSLGLGMSARLALIGVEGQPADGKELVSPCLSP 300

Db 217 LEQTPRGYLTFEMENSTYKLYTHSYLGFGLKAAKLATLGLAETE-GTDGHTFRSACLPR 275

QY 301 SFKREMEHAETTVYVSGQKAAASHLHCAARVSEVLQNRVHRTEYKAVDFAFSYYDL 360

Db 276 WLEAEWIFGVKYQYGGNQBSEVGPPEPCYAEVLNVRVSKLHQPEEVORGSFYAFSYYDR 335

QY 361 AAGVGLIDAEKGSILVVGDFEIAKVCRTLETQPSQSPSCMDLTYSLLQOE-FGFPFR 419

Db 336 AVDTDMIDYEKGGLTKVEDFERKAREVCNDLNEENTSSPFLCMDLSYITALLDGSFGFAD 395

QY 420 SKVTLTKRKIDNVETSMALGALFYHYDSL 448

Db 396 STVLTQTKKVNNIETGMALGATFHLQSL 424

RESULT 12

AAB72243

ID AAB72243 standard; protein; 428 AA.

XX AAB72243;

XX 14-MAY-2001 (first entry)

Human CD39 like protein CD39-L4 amino acid sequence.

Human CD39-like protein; apyrase; NDase; platelet function inhibitor;

myocardial infarction; cerebral ischaemia; angina; arterial thrombosis;

cerebral artery thrombosis; platelet aggregation; inflammation;

apoptosis; autoimmune disorder; neurological disorder;

Alzheimer's disease; Parkinson's disease; cancer; CD39-L4.

Homo sapiens.

MO200110205-A1.

15-FEB-2001.

09-AUG-2000; 2000MO-US021790.

09-AUG-1999; 99US-00370265.

11-JAN-2000; 2000US-00481238.

25-APR-2000; 2000US-00557800.

26-MAY-2000; 2000US-00583231.

30-JUN-2000; 2000US-00608285.



CC or chemokine-induced lung injury, inflammatory bowel disease or Crohn's  
disease. The present sequence is human CD39L4 protein

XX Sequence 428 AA;

Query Match 42.3%; Score 999; DB 5; Length 428;  
Best Local Similarity 52.4%; Pred. No. 5.5e-86;  
Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;

```
QY 62 PLGTAAAGHEVYFIMDAGSTGRVAVFOPT-RPPREFTLTHETFAVKYRGLSAYDD 120
DB 40 PINVSA---STLYGIMFDAGSTGRVAVFOPT-QMPQGLILEGEVFDVSPGLSAFVDQ 96
QY 121 VEKSAOGIRELIDVAKODIPDFWKATPVLKATAGLRLPGKAKQLQKVEVFAKSP 180
DB 97 PKQGAETVQGLLEVAKOSIRSHWKTTPVLKATAGLRLPGKAKQLLEVEKIFRKSP 156
QY 181 FLVDDCVSINMGTEGVSAWITINFLTGSILKTPGSSVGMLDGGSIOIALPRVEGT 240
DB 157 FLVFKGSVSIWDSDEGLAMVTNFTLQGLHGRQETVGLDGGASTOITFLPQPEKT 216
QY 241 LQASPFGVLTALRMNRTYKLYSYGLGMSARLAILGGVEGQPAKDKELVSPCLSP 300
DB 217 LEQTPRGVLTSEFNSTYKLYTSYGLFKAKRLATLGAETE-GTDGHTFRSAQLPR 275
QY 301 SFKGEWEHAETVYRVSGQKAAALHDLCAARVSEVLQNRVHTEEVKAVDFYAFSYDDL 360
DB 276 WLEAEWIFGVYKQYGGNGEGVEFPCYAEVLVVRGKLHQPPEVGRGSFYAFSYYYDR 335
QY 361 AAGVGLIDAEKGSIVNGDFEIAKYCRITETPOGSSPCMDLTVYSLLQE-RGFPFR 419
DB 336 AVDTDMIDYKGGILKVEDFERKAEVCDNENFTSGSPFLCMDLSTITALLKDGFGFAD 395
```

RESULT 14  
AAY44850  
ID AAY44850 standard; protein; 428 AA.

XX AAY44850;  
XX AC  
XX 18-MAY-2000 (first entry)

XX Human CD39-L4 variant-ACR III mutant protein.

XX CD39-L4; human; apyrase; nucleotide diphosphatase; NDPase; variant;  
XX ATP diphosphohydrolase; ADPbase; adenosine diphosphate; ADP; treatment;  
XX platelet aggregation; antithrombotic; thrombosis; myocardial infarction;  
XX cerebral ischaemia; angina; vascular graft; extracorporeal circulation;  
XX molecular weight marker; nutritional supplement; tumour prevention;  
XX drug targeting; substitution mutation.

XX Homo sapiens.  
XX Synthetic.

XX Key Location/Qualifiers  
XX FT Misc-difference 168 /note= "Wild type Asp substituted with Thr"  
XX FT Misc-difference 170 /note= "Wild type Ser substituted with Gln"  
XX FT Misc-difference 175 /note= "Wild type Leu substituted with Phe"  
XX FT

XX WO200004041-A2.

XX 27-JAN-2000.

XX 16-JUL-1999; 99WO-US016180.

XX 16-JUL-1998; 98US-00118205.

PR 24-JUL-1998; 98US-00122449.  
PR 04-FEB-1999; 99US-00244444.  
PR 19-MAR-1999; 99US-00273447.  
PR 09-JUL-1999; 99US-00350836.

XX (HYSE-) HYSEQ INC.

XX Ford J, Mulero J;

XX WPI: 2000-182397/16.

XX N-PGDB; AA250357.

XX New nucleic acid encoding human CD39-L4 protein, useful for treating  
and preventing thrombotic disease.

XX Claim 17; Fig 6; 125pp; English.

CC The present amino acid sequence is the CD39-L4 variant, designated as ACR  
CC III mutant protein, an apyrase and/or nucleotide diphosphatase (NDPase).  
CC It is isolated from the human foetal liver-spleen cDNA library,  
CC b2HFS20W. It is a soluble ATP diphosphohydrolase (ADPase) and is  
CC involved in the hydrolysis of adenosine diphosphate (ADP), the agonist  
CC that causes platelet aggregation. CD39-L4 protein has 30% and 80%  
CC homology to human and murine CD39. It has platelet aggregation inhibition  
CC and antithrombotic activity. CD39-L4 is used to treat or prevent  
CC thrombosis, myocardial infarction, cerebral ischaemia and angina. It is  
CC also used in vitro, to maintain vascular grafts or during extracorporeal  
CC circulation, to hydrolyse NDP, as molecular weight markers and as  
CC nutritional supplements. It is used to identify therapeutic agents that  
CC bind and modulate CD39-L4. It is coupled to toxins for targeting drugs to  
CC tumours or other cells that express CD39-L4

XX Sequence 428 AA;

Query Match 42.1%; Score 996; DB 3; Length 428;  
Best Local Similarity 52.4%; Pred. No. 1.1e-85;  
Matches 204; Conservative 54; Mismatches 125; Indels 6; Gaps 4;

```
QY 62 PLGTAAAGHEVYFIMDAGSTGRVAVFOPT-RPPREFTLTHETFAVKYRGLSAYDD 120
DB 40 PINVSA---STLYGIMFDAGSTGRVAVFOPT-QMPQGLILEGEVFDVSPGLSAFVDQ 96
QY 121 VEKSAOGIRELIDVAKODIPDFWKATPVLKATAGLRLPGKAKQLQKVEVFAKSP 180
DB 97 PKQGAETVQGLLEVAKOSIRSHWKTTPVLKATAGLRLPGKAKQLLEVEKIFRKSP 156
QY 181 FLVDDCVSINMGTEGVSAWITINFLTGSILKTPGSSVGMLDGGSIOIALPRVEGT 240
DB 157 FLVFKGSVSIWDSDEGLAMVTNFTLQGLHGRQETVGLDGGASTOITFLPQPEKT 216
QY 241 LQASPFGVLTALRMNRTYKLYSYGLGMSARLAILGGVEGQPAKDKELVSPCLSP 300
DB 217 LEQTPRGVLTSEFNSTYKLYTSYGLFKAKRLATLGAETE-GTDGHTFRSAQLPR 275
QY 301 SFKGEWEHAETVYRVSGQKAAALHDLCAARVSEVLQNRVHTEEVKAVDFYAFSYDDL 360
DB 276 WLEAEWIFGVYKQYGGNGEGVEFPCYAEVLVVRGKLHQPPEVGRGSFYAFSYYYDR 335
QY 361 AAGVGLIDAEKGSIVNGDFEIAKYCRITETPOGSSPCMDLTVYSLLQE-RGFPFR 419
DB 336 AVDTDMIDYKGGILKVEDFERKAEVCDNENFTSGSPFLCMDLSTITALLKDGFGFAD 395
```

RESULT 15  
AAB72240  
ID AAB72240 standard; protein; 428 AA.

XX AAB72240;



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 09:50:40 ; Search time 49 Seconds  
(without alignments)  
2618.172 Million cell updates/sec

Title: US-09-905-589A-2

Perfect score: 2364

Sequence: 1 MRKISNHSURVAKVAVPLG.....ALGATRYIDSLNRQSPAS 456

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 28138677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2364	100.0	484	13	US-10-092-063-27
2	2364	100.0	484	14	US-10-286-926-27
3	2364	100.0	484	15	US-10-231-913-123
4	2361	99.9	467	15	US-10-231-913-36
5	2357	99.7	484	15	US-10-231-913-124
6	2116	89.5	446	15	US-10-231-913-125
7	2063.5	87.3	455	15	US-10-231-913-125
8	1954	82.7	379	15	US-10-231-913-271
9	1007	42.6	427	15	US-10-231-913-126
10	999	42.3	428	13	US-10-091-085-3
11	999	42.3	428	13	US-10-091-085-3
12	999	42.3	428	13	US-10-092-063-3
13	999	42.3	428	13	US-10-092-063-5
14	999	42.3	428	14	US-10-286-926-3
15	999	42.3	428	14	US-10-286-926-5

16	999	42.3	428	15	US-10-231-913-127
17	996	42.1	428	13	US-10-091-085-7
18	996	42.1	428	13	US-10-092-063-7
19	996	42.1	428	14	US-10-286-926-7
20	992	38.5	428	16	US-10-408-765A-2296
21	909	38.5	405	13	US-10-092-063-25
22	909	38.5	405	14	US-10-286-926-25
23	909	38.5	465	13	US-10-092-063-39
24	823	34.8	330	9	US-09-925-299-876
25	823	34.8	330	10	US-09-925-299-876
26	590.5	25.0	479	15	US-10-369-493-6447
27	525.5	22.2	467	9	US-09-129-112-19
28	508.5	21.5	433	12	US-10-425-114-26762
29	508	21.5	556	15	US-10-369-493-1169
30	506	21.4	462	9	US-09-129-112-2
31	504	21.3	467	12	US-10-425-114-45875
32	501	21.2	459	9	US-09-129-112-9
33	500	21.2	467	12	US-10-424-599-230158
34	498	21.1	518	15	US-10-369-493-1713
35	490.5	20.7	443	12	US-10-425-114-49933
36	483	20.4	457	12	US-10-425-114-51762
37	483	20.4	459	12	US-10-424-599-145076
38	478.5	20.2	410	15	US-10-231-913-272
39	463	19.6	462	9	US-09-129-112-15
40	447	18.9	465	12	US-10-425-114-37241
41	447	18.9	465	12	US-10-425-114-49932
42	445	18.8	605	15	US-10-369-493-3333
43	443	18.7	455	14	US-10-259-165-786
44	392	16.6	459	14	US-10-259-165-74
45	380.5	16.1	540	12	US-10-424-599-283225

#### ALIGNMENTS

RESULT 1  
US-10-092-063-27  
Sequence 27, Application US/10092063  
Publication No. US20020173005A1  
GENERAL INFORMATION:  
APPLICANT: Maleco, Thilo  
TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE POLYPEPTIDES  
FILE REFERENCE: 28110735908  
CURRENT APPLICATION NUMBER: US/10/092, 063  
CURRENT FILING DATE: 2002-03-05  
PRIOR APPLICATION NUMBER: 09/370, 265  
PRIOR FILING DATE: 2002-01-31  
PRIOR APPLICATION NUMBER: PCT/US99/16180  
PRIOR FILING DATE: 1999-07-16  
PRIOR APPLICATION NUMBER: 09/350, 836  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: 09/273, 447  
PRIOR FILING DATE: 1999-03-19  
PRIOR APPLICATION NUMBER: 09/244, 444  
PRIOR FILING DATE: 1999-02-04  
PRIOR APPLICATION NUMBER: 09/122, 449  
PRIOR FILING DATE: 1998-07-24  
PRIOR APPLICATION NUMBER: 09/118, 205  
PRIOR FILING DATE: 1998-07-16  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 27  
LENGTH: 464  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-092-063-27

Query Match 100.0%; Score 2364; DB 13; Length 484;  
Best Local Similarity 100.0%; Pred. No. 1.4e-231;  
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 1 MRKISNHSURVAKVAVPLGCVGVFIYVAYIKMRATATQAFSITRAFGARWQQA 60

Db 29 MRKISNHSGLRVAKVAYPLGLCVGVFIYVAYIKWHRATATQAFFSITRAAPGARWGOAH 88  
QY 61 SPLGTAADGHEVYGVIMFDAGSGTGRVHVQFTRPPRETPTLTHETFKAVKGLSAYADD 120  
Db 89 SPLGTAADGHEVYGVIMFDAGSGTGRVHVQFTRPPRETPTLTHETFKAVKGLSAYADD 148  
QY 121 VEKSAQGIREDLLVAKQDIPFDWKATPLVLKATAGRLLLPGEKAQKLQKVEVFKASP 180  
Db 149 VEKSAQGIREDLLVAKQDIPFDWKATPLVLKATAGRLLLPGEKAQKLQKVEVFKASP 208  
QY 181 FLVGDDCVSVMNGTDEGVSAWITINFLTGSLKTPGGSSVGMDDLGGSGTQIAFLPRVEGT 240  
Db 209 FLVGDDCVSVMNGTDEGVSAWITINFLTGSLKTPGGSSVGMDDLGGSGTQIAFLPRVEGT 268  
QY 241 LQASPPGYLTALRMFNRTYKLYSYSLGGLMSARLAILGGVEGQPAKDGKELVSPCLSP 300  
Db 269 LQASPPGYLTALRMFNRTYKLYSYSLGGLMSARLAILGGVEGQPAKDGKELVSPCLSP 328  
QY 301 SPKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTVEVKHVDYFAFSYYIDL 360  
Db 329 SPKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTVEVKHVDYFAFSYYIDL 388  
QY 361 AAGVGLIDAEGKGLVVGDFEIAAKYVCRILETQPOSSPFCMDLTYVSLLLQEGFPPRS 420  
Db 389 AAGVGLIDAEGKGLVVGDFEIAAKYVCRILETQPOSSPFCMDLTYVSLLLQEGFPPRS 448  
QY 421 KVLKTRKIDNVETSWALGAIFHYIDSLNRQKSPAS 456  
Db 449 KVLKTRKIDNVETSWALGAIFHYIDSLNRQKSPAS 484

## RESULT 2

US-10-286-926-27  
; Sequence 27, Application US/10286926  
; Publication No. US2003017552A1  
; GENERAL INFORMATION:  
; APPLICANT: Ford, John  
; APPLICANT: Mulero, Julio  
; APPLICANT: Yeung, George  
; TITLE OF INVENTION: Methods and Materials Relating to CD39-Like  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 28110/36457CON  
; CURRENT APPLICATION NUMBER: US/10/286,926  
; CURRENT FILING DATE: 2002-11-01  
; PRIOR APPLICATION NUMBER: 09/557,800  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/481,238  
; PRIOR FILING DATE: 2000-01-11  
; PRIOR APPLICATION NUMBER: 09/370,265  
; PRIOR FILING DATE: 1999-08-09  
; PRIOR APPLICATION NUMBER: PCT/US99/16180  
; PRIOR FILING DATE: 1999-07-16  
; PRIOR APPLICATION NUMBER: 09/350836  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: 09/273447  
; PRIOR FILING DATE: 1999-03-19  
; PRIOR APPLICATION NUMBER: 09/122449  
; PRIOR FILING DATE: 1998-07-24  
; PRIOR APPLICATION NUMBER: 09/244444  
; PRIOR FILING DATE: 1999-02-04  
; PRIOR APPLICATION NUMBER: 09/118,205  
; PRIOR FILING DATE: 1998-07-16  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 27  
; LENGTH: 484  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-286-926-27

Query Match 100.0%; Score 2364; DB 14; Length 484;  
Best Local Similarity 100.0%; Pred. No. 1.4e-231;

Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MRKISNHSGLRVAKVAYPLGLCVGVFIYVAYIKWHRATATQAFFSITRAAPGARWGOAH 60  
Db 29 MRKISNHSGLRVAKVAYPLGLCVGVFIYVAYIKWHRATATQAFFSITRAAPGARWGOAH 88  
QY 61 SPLGTAADGHEVYGVIMFDAGSGTGRVHVQFTRPPRETPTLTHETFKAVKGLSAYADD 120  
Db 89 SPLGTAADGHEVYGVIMFDAGSGTGRVHVQFTRPPRETPTLTHETFKAVKGLSAYADD 148  
QY 121 VEKSAQGIREDLLVAKQDIPFDWKATPLVLKATAGRLLLPGEKAQKLQKVEVFKASP 180  
Db 149 VEKSAQGIREDLLVAKQDIPFDWKATPLVLKATAGRLLLPGEKAQKLQKVEVFKASP 208  
QY 181 FLVGDDCVSVMNGTDEGVSAWITINFLTGSLKTPGGSSVGMDDLGGSGTQIAFLPRVEGT 240  
Db 209 FLVGDDCVSVMNGTDEGVSAWITINFLTGSLKTPGGSSVGMDDLGGSGTQIAFLPRVEGT 268  
QY 241 LQASPPGYLTALRMFNRTYKLYSYSLGGLMSARLAILGGVEGQPAKDGKELVSPCLSP 300  
Db 269 LQASPPGYLTALRMFNRTYKLYSYSLGGLMSARLAILGGVEGQPAKDGKELVSPCLSP 328  
QY 301 SPKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTVEVKHVDYFAFSYYIDL 360  
Db 329 SPKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTVEVKHVDYFAFSYYIDL 388  
QY 361 AAGVGLIDAEGKGLVVGDFEIAAKYVCRILETQPOSSPFCMDLTYVSLLLQEGFPPRS 420  
Db 389 AAGVGLIDAEGKGLVVGDFEIAAKYVCRILETQPOSSPFCMDLTYVSLLLQEGFPPRS 448  
QY 421 KVLKTRKIDNVETSWALGAIFHYIDSLNRQKSPAS 456  
Db 449 KVLKTRKIDNVETSWALGAIFHYIDSLNRQKSPAS 484

## RESULT 3

US-10-231-913-123  
; Sequence 123, Application US/10231913  
; Publication No. US20040005576A1  
; GENERAL INFORMATION:  
; APPLICANT: Guo, Xiaojia S.  
; APPLICANT: Li, Li  
; APPLICANT: Fatturajan, Meera  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Casman, Stacie J.  
; APPLICANT: Malyankar, Uriel M.  
; APPLICANT: Tchernev, Velizar T.  
; APPLICANT: Vernet, Corine A.  
; APPLICANT: Spytek, Kimberly A.  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Alsobrook II, John P.  
; APPLICANT: Edinger, Schlomit  
; APPLICANT: Peyman, John A.  
; APPLICANT: Stone, David J.  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Gangolli, Esha A.  
; APPLICANT: Boldog, Ference L.  
; APPLICANT: Colman, Steven D.  
; APPLICANT: Eisen, Andrew J.  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Spaderna, Steven K.  
; APPLICANT: Zethusen, Bryan D.  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-216  
; CURRENT APPLICATION NUMBER: US/10/231,913  
; CURRENT FILING DATE: 2002-08-30  
; PRIOR APPLICATION NUMBER: 60/251,660  
; PRIOR FILING DATE: 2000-12-06  
; PRIOR APPLICATION NUMBER: 60/255,029  
; PRIOR FILING DATE: 2000-12-12  
; PRIOR APPLICATION NUMBER: 60/260,326  
; PRIOR FILING DATE: 2001-01-08

PRIOR APPLICATION NUMBER: 60/263,800  
PRIOR FILING DATE: 2001-01-24  
PRIOR APPLICATION NUMBER: 60/269,942  
PRIOR FILING DATE: 2001-02-20  
PRIOR APPLICATION NUMBER: 60/286,183  
PRIOR FILING DATE: 2001-04-24  
PRIOR APPLICATION NUMBER: 60/313,627  
PRIOR FILING DATE: 2001-08-20  
PRIOR APPLICATION NUMBER: 60/318,712  
PRIOR FILING DATE: 2001-09-12  
NUMBER OF SEQ ID NOS: 292  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO: 123  
LENGTH: 484  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-231-913-123

Query Match 100.0%; Score 2364; DB 15; Length 484;  
Best Local Similarity 100.0%; Pred. No. 1,4e-231;  
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MRKISNHSGLRVAKAVYPLGCVGFYVAYIKMHRATATQAFSITRAAPGARWGQAH 60  
29 MRKISNHSGLRVAKAVYPLGCVGFYVAYIKMHRATATQAFSITRAAPGARWGQAH 88  
61 SPLGTADGHEVFGYIMPDAGSTGRVHVFQTRPREPTLTHTFKAVKPGLSAYAD 120  
89 SPLGTADGHEVFGYIMPDAGSTGRVHVFQTRPREPTLTHTFKAVKPGLSAYAD 148  
121 VEKSAQGIKRELLDVAKODIPDFWKATPLVKATAGRLPPEKAKQLQKYEVEFKASP 180  
149 VEKSAQGIKRELLDVAKODIPDFWKATPLVKATAGRLPPEKAKQLQKYEVEFKASP 208  
181 FLVGGDDCVSINMGTDGVSAMITINFLGSLKTPGSSVGMLDIGGSTQIAFLPRVEGT 240  
209 FLVGGDDCVSINMGTDGVSAMITINFLGSLKTPGSSVGMLDIGGSTQIAFLPRVEGT 268  
241 LQASPPGYLTALRMFRYTKLYSYSLGLGMSARLALIGVEGQPAKDGKELVSPCLSP 300  
269 LQASPPGYLTALRMFRYTKLYSYSLGLGMSARLALIGVEGQPAKDGKELVSPCLSP 328  
301 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHTEEVKVDYFAFSYYDL 360  
329 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHTEEVKVDYFAFSYYDL 388  
361 AAGVGLIDAEKGSIVGDPEFAIKYVCTLETPOSSPFSQMDLTVVSLLLQEFGRPS 420  
389 AAGVGLIDAEKGSIVGDPEFAIKYVCTLETPOSSPFSQMDLTVVSLLLQEFGRPS 448  
421 KVLKTRKIDNVETSMALGAIFFHYIDSINRQKSPAS 456  
449 KVLKTRKIDNVETSMALGAIFFHYIDSINRQKSPAS 484

RESULT 4  
US-10-231-913-36  
Sequence 36, Application US/10231913  
Publication No. US20040005576A1  
GENERAL INFORMATION:  
APPLICANT: Guo, Xiaojia S.  
APPLICANT: Li, Li  
APPLICANT: Patnureajan, Meera  
APPLICANT: Shimkets, Richard A.  
APPLICANT: Casman, Stacie J.  
APPLICANT: Malvankar, Uriel M.  
APPLICANT: Tchernev, Velizar T.  
APPLICANT: Vernece, Corine A.  
APPLICANT: Spytek, Kimberly A.  
APPLICANT: Shenoy, Suresh G.  
APPLICANT: Alsbrook II, John P.  
APPLICANT: Edinger, Schlomit  
APPLICANT: Peyman, John A.

APPLICANT: Stone, David J.  
APPLICANT: Ellerman, Karen  
APPLICANT: Gangoli, Esna A.  
APPLICANT: Boldog, Ference L.  
APPLICANT: Coleman, Steven D.  
APPLICANT: Eileen, Andrew J.  
APPLICANT: Liu, Xiaohong  
APPLICANT: Padigaru, Muralidhara  
APPLICANT: Spaderna, Steven K.  
APPLICANT: Zernhusen, Bryan D.  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-216  
CURRENT APPLICATION NUMBER: US/10/231,913  
CURRENT FILING DATE: 2002-08-30  
PRIOR APPLICATION NUMBER: 60/251,660  
PRIOR FILING DATE: 2000-12-06  
PRIOR APPLICATION NUMBER: 60/255,029  
PRIOR FILING DATE: 2000-12-12  
PRIOR APPLICATION NUMBER: 60/260,326  
PRIOR FILING DATE: 2001-01-08  
PRIOR APPLICATION NUMBER: 60/263,800  
PRIOR FILING DATE: 2001-01-24  
PRIOR APPLICATION NUMBER: 60/269,942  
PRIOR FILING DATE: 2001-02-20  
PRIOR APPLICATION NUMBER: 60/286,183  
PRIOR FILING DATE: 2001-04-24  
PRIOR APPLICATION NUMBER: 60/313,627  
PRIOR FILING DATE: 2001-08-20  
PRIOR APPLICATION NUMBER: 60/318,712  
PRIOR FILING DATE: 2001-09-12  
NUMBER OF SEQ ID NOS: 292  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO: 367  
LENGTH: 467  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-231-913-36

Query Match 99.9%; Score 2361; DB 15; Length 467;  
Best Local Similarity 99.8%; Pred. No. 2,6e-231;  
Matches 455; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 MRKISNHSGLRVAKAVYPLGCVGFYVAYIKMHRATATQAFSITRAAPGARWGQAH 60  
12 MRKISNHSGLRVAKAVYPLGCVGFYVAYIKMHRATATQAFSITRAAPGARWGQAH 71  
61 SPLGTADGHEVFGYIMPDAGSTGRVHVFQTRPREPTLTHTFKAVKPGLSAYAD 120  
72 SPLGTADGHEVFGYIMPDAGSTGRVHVFQTRPREPTLTHTFKAVKPGLSAYAD 131  
121 VEKSAQGIKRELLDVAKODIPDFWKATPLVKATAGRLPPEKAKQLQKYEVEFKASP 180  
132 VEKSAQGIKRELLDVAKODIPDFWKATPLVKATAGRLPPEKAKQLQKYEVEFKASP 191  
181 FLVGGDDCVSINMGTDGVSAMITINFLGSLKTPGSSVGMLDIGGSTQIAFLPRVEGT 240  
192 FLVGGDDCVSINMGTDGVSAMITINFLGSLKTPGSSVGMLDIGGSTQIAFLPRVEGT 251  
241 LQASPPGYLTALRMFRYTKLYSYSLGLGMSARLALIGVEGQPAKDGKELVSPCLSP 300  
252 LQASPPGYLTALRMFRYTKLYSYSLGLGMSARLALIGVEGQPAKDGKELVSPCLSP 311  
301 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHTEEVKVDYFAFSYYDL 360  
312 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHTEEVKVDYFAFSYYDL 371  
361 AAGVGLIDAEKGSIVGDPEFAIKYVCTLETPOSSPFSQMDLTVVSLLLQEFGRPS 420  
372 AAGVGLIDAEKGSIVGDPEFAIKYVCTLETPOSSPFSQMDLTVVSLLLQEFGRPS 431  
421 KVLKTRKIDNVETSMALGAIFFHYIDSINRQKSPAS 456  
432 KVLKTRKIDNVETSMALGAIFFHYIDSINRQKSPAS 467

```
RESULT 5
US-10-231-913-124
; Sequence 124, Application US/10231913
; Publication No. US20040005576A1
; GENERAL INFORMATION:
; APPLICANT: Guo, Xiaojia S.
; APPLICANT: Li, Li
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Edinger, Schloimit
; APPLICANT: Peyman, John A.
; APPLICANT: Stone, David J.
; APPLICANT: Ellerman, Karen
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Boldog, Ference L.
; APPLICANT: Colman, Steven D.
; APPLICANT: Eisen, Andrew J.
; APPLICANT: Liu, Xiaohong
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Zehrusen, Bryan D.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-216
; CURRENT APPLICATION NUMBER: US/10/231,913
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: 60/251,660
; PRIOR FILING DATE: 2000-12-06
; PRIOR APPLICATION NUMBER: 60/255,029
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: 60/260,326
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: 60/263,800
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: 60/269,942
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/318,712
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 124
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-231-913-124

Query Match          99.78; Score 2357; DB 15; Length 484;
Best Local Similarity 99.68; Pred. No. 7e-231;
Matches 454; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKTISNHSRLRVAKVAYPLGLCVGVFTYVAYIKWHRATATQAFSTTRAAPGARWQQAH 60
Db 29 MKTISNHSRLRVAKVAYPLGLCVGVFTYVAYIKWHRATATQAFSTTRAAPGARWQQAH 88

Qy 61 SPLGTADAGHEVFGIMFDAGSTGTRVHVQFTRPPEPTLTTHETFKAVKGLSAYADD 120
Db 89 SPLGTADAGHEVFGIMFDAGSTGTRVHVQFTRPPEPTLTTHETFKAVKGLSAYADD 148

Qy 121 VEKSAQGIREDLLDVAKQDIFDFEWKATPLVLKATAGLRLLPGKAKLQKVKVEFKASP 180
Db 149 VEKSAQGIREDLLDVAKQDIFDFEWKATPLVLKATAGLRLLPGKAKLQKVKVEFKASP 208

Qy 181 FLVGDDCVSIMNGTDEGVSAMITINFLTGSLKTPGGSSVGMGLDLGGGSTQIAFLPRVEGT 240
Db 209 FLVGDDCVSIMNGTDEGVSAMITINFLTGSLKTPGGSSVGMGLDLGGGSTQIAFLPRVEGT 268
Qy 241 LQASPPGYLTALRMFNRTYKLYSYVGLGIMSGARLAILGGVEGQPAKDGKELVSPCLSP 300
Db 269 LQASPPGYLTALRMFNRTYKLYSYVGLGIMSGARLAILGGVEGQPAKDGKELVSPCLSP 328
Qy 301 SPKGWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTTEEVGHVDFYAFSYVDL 360
Db 329 SPKGWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTTEEVGHVDFYAFSYVDL 388
Qy 361 AAGVGLIDAEKGGSLVVGDFEIIAAKYVCRTELETOPOSSPSCMDLTYVSLLLQEFEGPERS 420
Db 389 AAGVGLIDAEKGGSLVVGDFEIIAAKYVCRTELETOPOSSPSCMDLTYVSLLLQEFEGPERS 448
Qy 421 KVLKLTREKIDNVETSWALGAIEHYIDSLNRQKSPAS 456
Db 449 KVLKLTREKIDNVETSWALGAIEHYIDSLNRQKSPAS 484

RESULT 6
US-10-231-913-38
; Sequence 38, Application US/10231913
; Publication No. US20040005576A1
; GENERAL INFORMATION:
; APPLICANT: Guo, Xiaojia S.
; APPLICANT: Li, Li
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Edinger, Schloimit
; APPLICANT: Peyman, John A.
; APPLICANT: Stone, David J.
; APPLICANT: Ellerman, Karen
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Boldog, Ference L.
; APPLICANT: Colman, Steven D.
; APPLICANT: Eisen, Andrew J.
; APPLICANT: Liu, Xiaohong
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Zehrusen, Bryan D.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-216
; CURRENT APPLICATION NUMBER: US/10/231,913
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: 60/251,660
; PRIOR FILING DATE: 2000-12-06
; PRIOR APPLICATION NUMBER: 60/255,029
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: 60/260,326
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: 60/263,800
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: 60/269,942
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/318,712
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 38
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
```



Tue Jun 8 12:10:38 2004

us-09-905-589a-2.rapb

Page 5

LENGTH: 446  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-231-913-38

Query Match 89.5%; Score 2116; DB 15; Length 446;  
Best Local Similarity 91.0%; Pred. No. 2,2e-206;  
Matches 415; Conservative 1; Mismatches 2; Indels 38; Gaps 1;

```
QY 1 MKTISNHSRVAKVAAPLGLCVGFIVAYIKMHRATATQAFSITRAAPAGAWGQAH 60
DB 29 MKTISNHSRVA--
QY 61 SPLGTADGHEVFYGMFDAGSTGTRVHVQFTRPREPTTLTHTETPKAVKGLSAYAD 120
DB 51 SPLGTADGHEVFYGMFDAGSTGTRVHVQFTRPREPTTLTHTETPKAVKGLSAYAD 110
QY 121 VEKSAQGIREFLLVAKODIFPDWKAATPLVLTAKATAGRLPGEKAKLQKVEVFKA 180
DB 111 VEKSAQGIREFLLVAKODIFPDWKAATPLVLTAKATAGRLPGEKAKLQKVEVFKA 170
QY 181 FLVGDCCVSIMNGTDEGVSAWITINPLTGLSKTPGSSVGMULDGGSTQIAFLPRE 240
DB 171 FLVGDCCVSIMNGTDEGVSAWITINPLTGLSKTPGSSVGMULDGGSTQIAFLPRE 230
QY 241 LQASPPGYLTALRMNRTYLYSYGLGLMSARLAILGGVFGOPAKGKELVSPCLSP 300
DB 231 LQASPPGYLTALRMNRTYLYSYGLGLMSARLAILGGVFGOPAKGKELVSPCLSP 290
QY 301 SFKGEWEHAETVTVRSQKAAASLHELCAARVSEVLQNRVHRTVEYKVDYFAFSYY 360
DB 291 SFKGEWEHAETVTVRSQKAAASLHELCAARVSEVLQNRVHRTVEYKVDYFAFSYY 350
QY 361 AAGVGLIDAEKGGSLVVDDEFIAKVCRTLETQPOSSPCMDLTYVSLILOEFGFP 420
DB 351 AAGVGLIDAEKGGSLVVDDEFIAKVCRTLETQPOSSPCMDLTYVSLILOEFGFP 410
QY 421 KVLKLTTRKIDIVETSMALGAIFFHYIDSLNROKSPAS 456
DB 411 KVLKLTTRKIDIVETSMALGAIFFHYIDSLNROKSPAS 446
```

RESULT 7  
US-10-231-913-125  
Sequence 125, Application US/10231913  
Publication No. US20040005576A1  
GENERAL INFORMATION:  
APPLICANT: Guo, Xiaojia S.  
APPLICANT: Li, Li  
APPLICANT: Patuturajan, Meera  
APPLICANT: Shinkets, Richard A.  
APPLICANT: Caeman, Stacie J.  
APPLICANT: Malvankar, Uriel M.  
APPLICANT: Tchenev, Velizar I.  
APPLICANT: Verneel, Corine A.  
APPLICANT: Spytek, Kimberly A.  
APPLICANT: Shenoy, Suresh G.  
APPLICANT: Alebrock II, John P.  
APPLICANT: Edinger, Schlomit  
APPLICANT: Peyman, John A.  
APPLICANT: Stone, David J.  
APPLICANT: Ellerman, Karen  
APPLICANT: Gangoli, Esha A.  
APPLICANT: Boldog, Ferenc L.  
APPLICANT: Colman, Steven D.  
APPLICANT: Eisen, Andrew J.  
APPLICANT: Liu, Xiaohong  
APPLICANT: Padigaru, Muralidhara  
APPLICANT: Spaderna, Steven K.  
APPLICANT: Zernusen, Bryan D.  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-216  
CURRENT APPLICATION NUMBER: US/10/231,913

CURRENT FILING DATE: 2002-08-30  
PRIOR APPLICATION NUMBER: 60/251,660  
PRIOR FILING DATE: 2000-12-06  
PRIOR APPLICATION NUMBER: 60/255,029  
PRIOR FILING DATE: 2000-12-12  
PRIOR APPLICATION NUMBER: 60/260,326  
PRIOR FILING DATE: 2001-01-08  
PRIOR APPLICATION NUMBER: 60/263,800  
PRIOR FILING DATE: 2001-01-24  
PRIOR APPLICATION NUMBER: 60/269,942  
PRIOR FILING DATE: 2001-02-20  
PRIOR APPLICATION NUMBER: 60/266,183  
PRIOR FILING DATE: 2001-04-24  
PRIOR APPLICATION NUMBER: 60/313,627  
PRIOR FILING DATE: 2001-08-20  
PRIOR APPLICATION NUMBER: 60/318,712  
PRIOR FILING DATE: 2001-09-12  
NUMBER OF SEQ ID NOS: 292  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 125  
LENGTH: 455  
TYPE: PRT  
ORGANISM: Rattus norvegicus  
US-10-231-913-125

Query Match 87.3%; Score 2063.5; DB 15; Length 455;  
Best Local Similarity 86.6%; Pred. No. 5.2e-201;  
Matches 394; Conservative 27; Mismatches 33; Indels 1; Gaps 1;

```
QY 1 MKTISNHSRVAKVAAPLGLCVGFIVAYIKMHRATATQAFSITRAAPAGAWGQAH 60
DB 1 MKTISNHSRVAKVAAPLGLCVGFIVAYIKMHRATATQAFSITRAAPAGAWGQAH 60
QY 61 SPLGTADGHEVFYGMFDAGSTGTRVHVQFTRPREPTTLTHTETPKAVKGLSAYAD 120
DB 61 SPLGTADGHEVFYGMFDAGSTGTRVHVQFTRPREPTTLTHTETPKAVKGLSAYAD 120
QY 121 VEKSAQGIREFLLVAKODIFPDWKAATPLVLTAKATAGRLPGEKAKLQKVEVFKA 180
DB 121 VEKSAQGIREFLLVAKODIFPDWKAATPLVLTAKATAGRLPGEKAKLQKVEVFKA 180
QY 181 FLVGDCCVSIMNGTDEGVSAWITINPLTGLSKTPGSSVGMULDGGSTQIAFLPRE 240
DB 181 FLVGDCCVSIMNGTDEGVSAWITINPLTGLSKTPGSSVGMULDGGSTQIAFLPRE 240
QY 241 LQASPPGYLTALRMNRTYLYSYGLGLMSARLAILGGVFGOPAKGKELVSPCLSP 300
DB 241 LQASPPGYLTALRMNRTYLYSYGLGLMSARLAILGGVFGOPAKGKELVSPCLSP 300
QY 301 SFKGEWEHAETVTVRSQKAAASLHELCAARVSEVLQNRVHRTVEYKVDYFAFSYY 360
DB 301 SFKGEWEHAETVTVRSQKAAASLHELCAARVSEVLQNRVHRTVEYKVDYFAFSYY 359
QY 361 AAGVGLIDAEKGGSLVVDDEFIAKVCRTLETQPOSSPCMDLTYVSLILOEFGFP 420
DB 360 AAGVGLIDAEKGGSLVVDDEFIAKVCRTLETQPOSSPCMDLTYVSLILOEFGFP 419
QY 421 KVLKLTTRKIDIVETSMALGAIFFHYIDSLNROKSPAS 455
DB 420 KVLKLTTRKIDIVETSMALGAIFFHYIDSLNROKSPAS 454
```

RESULT 8  
US-10-231-913-271  
Sequence 271, Application US/10231913  
Publication No. US20040005576A1  
GENERAL INFORMATION:  
APPLICANT: Guo, Xiaojia S.  
APPLICANT: Li, Li  
APPLICANT: Patuturajan, Meera  
APPLICANT: Shinkets, Richard A.  
APPLICANT: Caeman, Stacie J.  
APPLICANT: Malvankar, Uriel M.

```

; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Edinger, Schlomit
; APPLICANT: Peyman, John A.
; APPLICANT: Stone, David J.
; APPLICANT: Ellerman, Karen
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Boldog, Ference L.
; APPLICANT: Colman, Steven D.
; APPLICANT: Eisen, Andrew J.
; APPLICANT: Liu, Xiaohong
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Zerhusen, Bryan D.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-216
; CURRENT APPLICATION NUMBER: US/10/231,913
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: 60/251,660
; PRIOR FILING DATE: 2000-12-06
; PRIOR APPLICATION NUMBER: 60/255,029
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: 60/260,326
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: 60/263,800
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: 60/269,942
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/286,183
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/313,627
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/318,712
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 271
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-231-913-271

Query Match      82.7%; Score 1954; DB 15; Length 379;
Best Local Similarity 99.7%; Pred. No. 5.6e-190;
Matches 378; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 71 EVFYGMFDAGSTGRVHVVFQFTRPPRETPTLTHTFKAVKPGLSAYADDVEKSAQGIRE 130
Db 1 EVFYGMFDAGSTGRVHVVFQFTRPPRETPTLTHTFKALKEGLSAYADDVEKSAQGIRE 60
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 131 LLDVAKQDIPFDKATPLVLKATAGLRLPGKAQKLLQKVKPKASPFVLGDDCVSI 190
Db 61 LLDVAKQDIPFDKATPLVLKATAGLRLPGKAQKLLQKVKPKASPFVLGDDCVSI 120
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 191 MNGTDEGVSAWITINFLTGSKTPGSSVGMPLDGGSTQIAFLPVEGTLOASPPGYLT 250
Db 121 MNGTDEGVSAWITINFLTGSKTPGSSVGMPLDGGSTQIAFLPVEGTLOASPPGYLT 180
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 251 ALRMFNRTKLYSVYGLGLGMSARLAILGGVGGQAKGKELVSPCLSPSPKGEWEHAE 310
Db 181 ALRMFNRTKLYSVYGLGLGMSARLAILGGVGGQAKGKELVSPCLSPSPKGEWEHAE 240
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 311 VTYRVSGQKAASLHELCAARVSEVLQNRVHRTEEVKRVDFYAFSYYYDLAAGVGLIDAE 370
Db 241 VTYRVSGQKAASLHELCAARVSEVLQNRVHRTEEVKRVDFYAFSYYYDLAAGVGLIDAE 300
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 371 KGSLLVGDPEIAAKYVCRTLETQPOSSPSCMDLTVVSLLLQEFPPSKYLKTRKID 430
Db 301 KGSLLVGDPEIAAKYVCRTLETQPOSSPSCMDLTVVSLLLQEFPPSKYLKTRKID 360
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Edinger, Schlomit
; APPLICANT: Peyman, John A.
; APPLICANT: Stone, David J.
; APPLICANT: Ellerman, Karen
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Boldog, Ference L.
; APPLICANT: Colman, Steven D.
; APPLICANT: Eisen, Andrew J.
; APPLICANT: Liu, Xiaohong
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Zerhusen, Bryan D.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-216
; CURRENT APPLICATION NUMBER: US/10/231,913
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: 60/251,660
; PRIOR FILING DATE: 2000-12-06
; PRIOR APPLICATION NUMBER: 60/255,029
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: 60/260,326
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: 60/263,800
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: 60/269,942
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/286,183
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/313,627
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/318,712
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 271
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-231-913-271

Query Match      42.6%; Score 1007; DB 15; Length 427;
Best Local Similarity 50.0%; Pred. No. 2.2e-93;
Matches 213; Conservative 62; Mismatches 133; Indels 18; Gaps 6;

Qy 37 ATATQAFSTTRAAPGA-----RWGQ-----QAHSPLGTADGHEVFGIMFDAGSTG 84
Db 2 ATSGAVFMILLIACVSTVFYREQQTWFEVGLSSMCPINVSAG---TFYGINFDAGSTG 58
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 85 TRVHVFOFT-RPPRETPTLTHTFKAVKPGLSAYADDVEKSAQIRELLLDVAKQDIPFD 143
Db 59 TRIHVITFVQTAGQLPFLGEIIFDSVKPGLSAFVDQPKQGAETVQELLEVAKDSIPRSH 118
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

; Sequence 126, Application US/10231913
; Publication No. US20040005576A1
; GENERAL INFORMATION:
; APPLICANT: Guo, Xiaojia S.
; APPLICANT: Li, Li
; APPLICANT: Patturajan, Meera
; APPLICANT: Shinkets, Richard A.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Edinger, Schlomit
; APPLICANT: Peyman, John A.
; APPLICANT: Stone, David J.
; APPLICANT: Ellerman, Karen
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Boldog, Ference L.
; APPLICANT: Colman, Steven D.
; APPLICANT: Eisen, Andrew J.
; APPLICANT: Liu, Xiaohong
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Zerhusen, Bryan D.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-216
; CURRENT APPLICATION NUMBER: US/10/231,913
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: 60/251,660
; PRIOR FILING DATE: 2000-12-06
; PRIOR APPLICATION NUMBER: 60/255,029
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: 60/260,326
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: 60/263,800
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: 60/269,942
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/286,183
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/313,627
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/318,712
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 126
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Mus musculus
; US-10-231-913-126
```

QY 144 WKATPLVTKATAGTLLPGEKAKLLOKKEVFKASPLVGDCCVSIINOTDEGSAMIT 203  
 DB 119 WERTPVVKATAGTLLPGEKAKLLOKKEVFKASPLVGDCCVSIINOTDEGSAMIT 178  
 QY 204 INFLTGSILKTPGSSVGMGLDGGSTQIAFLPRVEGTLQASPPGYLTALMFRNRYKYS 263  
 DB 179 VNFITGLDHRGCGTGTGTLDDGASQTITFLPQEKTELEOTPRGYLTALMFRNRYKYS 238  
 QY 264 YSYIGLGMARLALIGCVSGQPKADGKELVPSLSPFGEMENHAYTVRVSGQAAAS 323  
 DB 239 HSYLGFGLKARLALIGCVSGQPKADGKELVPSLSPFGEMENHAYTVRVSGQAAAS 297  
 QY 324 LHEICARVSEVLONRVHRTTEVGVDFVAFSYYYDLAAGVLIDAEKGSILVGDPELA 383  
 DB 298 GFEPYAEVLRYVQKHLHQBEGVGSFVAFSYYYDLAAGVLIDAEKGSILVGDPELA 357  
 QY 384 AKYVCRLETOPGSSPSCMDLTVYSLLOE-FGPPSKVLKTRKIDNVEISMGALATF 442  
 DB 358 AREVCDMLGSSSGSPFLCMDLTVYSLLOE-FGPPSKVLKTRKIDNVEISMGALATF 417  
 QY 443 HYIDSL 448  
 DB 418 HLQSL 423

# RESULT 10

US-10-091-085-3  
 / Sequence 3, Application US/10091085  
 / Publication No. US20020146772A1  
 / GENERAL INFORMATION:  
 / APPLICANT: Mulero, Jullio  
 / TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE  
 / FILE REFERENCE: 28110/35761  
 / CURRENT APPLICATION NUMBER: US/10/091,085  
 / PRIOR FILING DATE: 2002-03-05  
 / PRIOR APPLICATION NUMBER: 09/350,836  
 / PRIOR FILING DATE: 1999-07-09  
 / PRIOR APPLICATION NUMBER: 09/273,447  
 / PRIOR FILING DATE: 1999-03-19  
 / PRIOR APPLICATION NUMBER: 09/118,205  
 / PRIOR FILING DATE: 1998-07-16  
 / PRIOR APPLICATION NUMBER: 09/122,449  
 / PRIOR FILING DATE: 1998-07-24  
 / PRIOR APPLICATION NUMBER: 09/244,444  
 / PRIOR FILING DATE: 1999-02-04  
 / NUMBER OF SEQ ID NOS: 23  
 / SOFTWARE: PatentIn Ver. 2.0  
 / SEQ ID NO 3  
 / LENGTH: 428  
 / TYPE: PRT  
 / ORGANISM: Homo sapiens  
 US-10-091-085-3

Query Match 42.3%; Score 999; DB 13; Length 428;  
 Best Local Similarity 52.4%; Pred. No. 1.4e-92;  
 Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;  
 QY 62 PLGTADGHEVYVYGMFDAGSTGRVHVFOPT-RPPREPTLTHETFAVKGSAVADD 120  
 DB 40 PINVSA---STLYGMFDAGSTGRVHVFOPT-RPPREPTLTHETFAVKGSAVADD 96  
 QY 121 VEKSAQIRELLDVAKODIPDFWKATPLVTKATAGTLLPGEKAKLLOKKEVFKASP 180  
 DB 97 PRQAGETVQGLLEVAKODIPDFWKATPLVTKATAGTLLPGEKAKLLOKKEVFKASP 156  
 QY 181 FLVGDCCVSIINOTDEGSAMITINFLTGSILKTPGSSVGMGLDGGSTQIAFLPRVEGT 240  
 DB 157 FLVPGKSVSIINOTDEGSAMITINFLTGSILKTPGSSVGMGLDGGSTQIAFLPRVEGT 216  
 QY 241 LQASPGVLTALMFRNRYKYSYVIGLGMARLALIGCVSGQPKADGKELVPSLSP 300

DB 217 LEQTPRGVLTSEFEMENSTYKLYTHSYLGFGLKARLALIGCVSGQPKADGKELVPSLSP 275  
 QY 301 SFKGEWENHAYTVRVSGQAAASLHEICARVSEVLONRVHRTTEVGVDFVAFSYYYDL 360  
 DB 276 WLEAEWIFGVYKQYQGGNQBEGVGFEPYAEVLRYVQKHLHQBEGVGSFVAFSYYYDL 335  
 QY 361 AAGVGLIDAEKGSILVGDPELAAYVCRLETOPGSSPSCMDLTVYSLLOE-FGPPR 419  
 DB 336 AVDTMIDYEKGLIKVDFERKAREVCDMLNENFTSGSPFLCMDLTVYSLLOE-FGPPR 395  
 QY 420 SKVLKTRKIDNVEISMGALATFHYIDSL 448  
 DB 396 STVLTQTKVNNIETGMALGATFHLQSL 424

# RESULT 11

US-10-091-085-5  
 / Sequence 5, Application US/10091085  
 / Publication No. US20020146772A1  
 / GENERAL INFORMATION:  
 / APPLICANT: Mulero, Jullio  
 / TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE  
 / FILE REFERENCE: 28110/35761  
 / CURRENT APPLICATION NUMBER: US/10/091,085  
 / PRIOR FILING DATE: 2002-03-05  
 / PRIOR APPLICATION NUMBER: 09/350,836  
 / PRIOR FILING DATE: 1999-07-09  
 / PRIOR APPLICATION NUMBER: 09/273,447  
 / PRIOR FILING DATE: 1999-03-19  
 / PRIOR APPLICATION NUMBER: 09/118,205  
 / PRIOR FILING DATE: 1998-07-16  
 / PRIOR APPLICATION NUMBER: 09/122,449  
 / PRIOR FILING DATE: 1998-07-24  
 / PRIOR APPLICATION NUMBER: 09/244,444  
 / PRIOR FILING DATE: 1999-02-04  
 / NUMBER OF SEQ ID NOS: 23  
 / SOFTWARE: PatentIn Ver. 2.0  
 / SEQ ID NO 5  
 / LENGTH: 428  
 / TYPE: PRT  
 / ORGANISM: Homo sapiens  
 US-10-091-085-5

Query Match 42.3%; Score 999; DB 13; Length 428;  
 Best Local Similarity 52.4%; Pred. No. 1.4e-92;  
 Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;  
 QY 62 PLGTADGHEVYVYGMFDAGSTGRVHVFOPT-RPPREPTLTHETFAVKGSAVADD 120  
 DB 40 PINVSA---STLYGMFDAGSTGRVHVFOPT-RPPREPTLTHETFAVKGSAVADD 96  
 QY 121 VEKSAQIRELLDVAKODIPDFWKATPLVTKATAGTLLPGEKAKLLOKKEVFKASP 180  
 DB 97 PRQAGETVQGLLEVAKODIPDFWKATPLVTKATAGTLLPGEKAKLLOKKEVFKASP 156  
 QY 181 FLVGDCCVSIINOTDEGSAMITINFLTGSILKTPGSSVGMGLDGGSTQIAFLPRVEGT 240  
 DB 157 FLVPGKSVSIINOTDEGSAMITINFLTGSILKTPGSSVGMGLDGGSTQIAFLPRVEGT 216  
 QY 241 LQASPGVLTALMFRNRYKYSYVIGLGMARLALIGCVSGQPKADGKELVPSLSP 300  
 DB 217 LEQTPRGVLTSEFEMENSTYKLYTHSYLGFGLKARLALIGCVSGQPKADGKELVPSLSP 275  
 QY 301 SFKGEWENHAYTVRVSGQAAASLHEICARVSEVLONRVHRTTEVGVDFVAFSYYYDL 360  
 DB 276 WLEAEWIFGVYKQYQGGNQBEGVGFEPYAEVLRYVQKHLHQBEGVGSFVAFSYYYDL 335  
 QY 361 AAGVGLIDAEKGSILVGDPELAAYVCRLETOPGSSPSCMDLTVYSLLOE-FGPPR 419  
 DB 336 AVDTMIDYEKGLIKVDFERKAREVCDMLNENFTSGSPFLCMDLTVYSLLOE-FGPPR 395

QY 420 SKVLKTRKIDNVETSWALGAIFHYIDSL 448  
Db 396 STVLQTKKNNIETGALGATFHLQSL 424

## RESULT 12

US-10-092-063-3

; Sequence 3, Application US/10092063

; Publication No. US20020173005A1

; GENERAL INFORMATION:

; APPLICANT: Ford, John

; APPLICANT: Mulero, Julio

; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE POLYPEPTIDES

; FILE REFERENCE: 28110/35908

; CURRENT APPLICATION NUMBER: US/10/092,063

; CURRENT FILING DATE: 2002-03-05

; PRIOR APPLICATION NUMBER: 09/370,265

; PRIOR FILING DATE: 2002-01-31

; PRIOR APPLICATION NUMBER: PCT/US99/16180

; PRIOR FILING DATE: 1999-07-16

; PRIOR APPLICATION NUMBER: 09/350,836

; PRIOR FILING DATE: 1999-07-09

; PRIOR APPLICATION NUMBER: 09/273,447

; PRIOR FILING DATE: 1999-03-19

; PRIOR APPLICATION NUMBER: 09/244,444

; PRIOR FILING DATE: 1999-02-04

; PRIOR APPLICATION NUMBER: 09/122,449

; PRIOR FILING DATE: 1998-07-24

; PRIOR APPLICATION NUMBER: 09/118,205

; PRIOR FILING DATE: 1998-07-16

; NUMBER OF SEQ ID NOS: 39

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3

; LENGTH: 428

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-092-063-3

Query Match 42.3%; Score 999; DB 13; Length 428;

Best Local Similarity 52.4%; Pred. No. 1.4e-92;

Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;

QY 62 PLGTAADGHEVFYGMFDAGSTGTRVHVFOQT-RPPRETPTLTHTETFKAVKPGLSAYADD 120  
Db 40 PINVSA---STLYGIMFDAGSTGTRVHVTFVQKMPGQLPILGEVDFSVKPGLSAFVDQ 96

QY 121 VEKSAQGIREDLVAKQDIPDFWKATPLVLKATAGRLLPGEKAKQLLQKVEVFKASP 180  
Db 97 PKQGAETVQGLLEAVKDSIPRSHWKKTPVVLKATAGRLLPGEKAKALLFEVKEIFRKSP 156

QY 181 FLVGDDCVSIMGTDGVSANITINFLTGLSKATPGSSVGMLDLGGSGTQIAFLPRVEGT 240  
Db 157 FLVPKGSVIMDSDEGILAWTVNFLTQQLHGHROETVGTLDLGGASTQITFLPQPEKT 216

QY 241 LQASPPGYLTALRMFNRTYKLYSYLGLMSARLAILGGVEGQPAKDGKELVSPCLSP 300  
Db 217 LEQTPRGYLTSPFMFNSTYKLYTHSYLGLKAAALATLGALETE-GTDGHTFRSACLPR 275

QY 301 SPKGEWEHAEVTVYRVSQQAASLHELCAARYSEVLONRVHRTVEVKHVDVFAFSYYIDL 360  
Db 276 WLEAEWIFGKVYQYGGNOGEVGEVPEPCYAEVLVRVVRGKLHQPEEVQSGSFYAFSYYDR 335

QY 361 AAGVGLIDAEKGGSLVVGDPFEIAAKVVCRTLETQPOSSPFCMDLTYVSLLIQOE-FGPPR 419  
Db 336 AVDTDMIDYKGGILKVEDFERKAREVCNLENFTSGSPFLCMLDSYITALLKDGFGFAD 395

QY 420 SKVLKTRKIDNVETSWALGAIFHYIDSL 448  
Db 396 STVLQTKKNNIETGALGATFHLQSL 424

## RESULT 13

US-10-092-063-5

; Sequence 5, Application US/10092063

; Publication No. US20020173005A1

; GENERAL INFORMATION:

; APPLICANT: Ford, John

; APPLICANT: Mulero, Julio

; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE POLYPEPTIDES

; FILE REFERENCE: 28110/35908

; CURRENT APPLICATION NUMBER: US/10/092,063

; CURRENT FILING DATE: 2002-03-05

; PRIOR APPLICATION NUMBER: 09/370,265

; PRIOR FILING DATE: 2002-01-31

; PRIOR APPLICATION NUMBER: PCT/US99/16180

; PRIOR FILING DATE: 1999-07-16

; PRIOR APPLICATION NUMBER: 09/350,836

; PRIOR FILING DATE: 1999-07-09

; PRIOR APPLICATION NUMBER: 09/273,447

; PRIOR FILING DATE: 1999-03-19

; PRIOR APPLICATION NUMBER: 09/244,444

; PRIOR FILING DATE: 1999-02-04

; PRIOR APPLICATION NUMBER: 09/122,449

; PRIOR FILING DATE: 1998-07-24

; PRIOR APPLICATION NUMBER: 09/118,205

; PRIOR FILING DATE: 1998-07-16

; NUMBER OF SEQ ID NOS: 39

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 5

; LENGTH: 428

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-092-063-5

Query Match 42.3%; Score 999; DB 13; Length 428;

Best Local Similarity 52.4%; Pred. No. 1.4e-92;

Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;

QY 62 PLGTAADGHEVFYGMFDAGSTGTRVHVFOQT-RPPRETPTLTHTETFKAVKPGLSAYADD 120  
Db 40 PINVSA---STLYGIMFDAGSTGTRVHVTFVQKMPGQLPILGEVDFSVKPGLSAFVDQ 96

QY 121 VEKSAQGIREDLVAKQDIPDFWKATPLVLKATAGRLLPGEKAKQLLQKVEVFKASP 180  
Db 97 PKQGAETVQGLLEAVKDSIPRSHWKKTPVVLKATAGRLLPGEKAKALLFEVKEIFRKSP 156

QY 181 FLVGDDCVSIMGTDGVSANITINFLTGLSKATPGSSVGMLDLGGSGTQIAFLPRVEGT 240  
Db 157 FLVPKGSVIMDSDEGILAWTVNFLTQQLHGHROETVGTLDLGGASTQITFLPQPEKT 216

QY 241 LQASPPGYLTALRMFNRTYKLYSYLGLMSARLAILGGVEGQPAKDGKELVSPCLSP 300  
Db 217 LEQTPRGYLTSPFMFNSTYKLYTHSYLGLKAAALATLGALETE-GTDGHTFRSACLPR 275

QY 301 SPKGEWEHAEVTVYRVSQQAASLHELCAARYSEVLONRVHRTVEVKHVDVFAFSYYIDL 360  
Db 276 WLEAEWIFGKVYQYGGNOGEVGEVPEPCYAEVLVRVVRGKLHQPEEVQSGSFYAFSYYDR 335

QY 361 AAGVGLIDAEKGGSLVVGDPFEIAAKVVCRTLETQPOSSPFCMDLTYVSLLIQOE-FGPPR 419  
Db 336 AVDTDMIDYKGGILKVEDFERKAREVCNLENFTSGSPFLCMLDSYITALLKDGFGFAD 395

QY 420 SKVLKTRKIDNVETSWALGAIFHYIDSL 448  
Db 396 STVLQTKKNNIETGALGATFHLQSL 424

## RESULT 14

US-10-286-926-3

; Sequence 3, Application US/10286926

; Publication No. US20030175752A1

; GENERAL INFORMATION:

; APPLICANT: Ford, John

; APPLICANT: Mulero, Julio

; APPLICANT: Yeung, George

; TITLE OF INVENTION: Methods and Materials Relating to CD39-Like

FILE OF INVENTION: Polypeptides  
FILE REFERENCE: 28110/36457CON  
CURRENT APPLICATION NUMBER: US/10/286,926  
CURRENT FILING DATE: 2002-11-01  
PRIOR APPLICATION NUMBER: 09/557,800  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/481,238  
PRIOR FILING DATE: 2000-01-11  
PRIOR APPLICATION NUMBER: 09/370,265  
PRIOR FILING DATE: 1999-08-09  
PRIOR APPLICATION NUMBER: PCT/US99/16180  
PRIOR FILING DATE: 1999-07-16  
PRIOR APPLICATION NUMBER: 09/350836  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: 09/273447  
PRIOR FILING DATE: 1999-03-19  
PRIOR APPLICATION NUMBER: 09/122449  
PRIOR FILING DATE: 1998-07-24  
PRIOR APPLICATION NUMBER: 09/244444  
PRIOR FILING DATE: 1999-02-04  
PRIOR APPLICATION NUMBER: 09/118,205  
PRIOR FILING DATE: 1998-07-16  
NUMBER OF SEQ ID NOS: 54  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 3  
LENGTH: 428  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-286-926-5

Query Match 42.3%; Score 999; DB 14; Length 428;  
Best Local Similarity 52.4%; Pred. No. 1.4e-92;  
Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;  
QY 62 PLGTADGHEVFFYGMFAGSTGTGVHVFQFT-RRPREPTLTHETFAVKXGLSAYADD 120  
DB 40 PIVNSA---STLYGIMFAGSTGTGRHIVYTFVQKMPGQLPILGGEVFDVSKGLSAYDQ 96  
QY 121 VEKSAQIRELLDVAKODIPDFFWKATPLVKATAGRLLPGEKAKOKLQKYKEVFKASP 180  
DB 97 PROGAEITVOGLLEVAKDSIPRSHMKKTPLVKATAGRLLPGEKAKALLFEVKEIFPKSP 156  
QY 161 FLVGGDCVSIANGTDEGVSAWITINFLTGLSGKTPGSSVGMULDGGSGTQIAFLPREVGT 240  
DB 157 FLVPGKSVSINDGSDGGLAWTVNFLTQGLHGRQETVGLDGGASTQITFLPOFEKT 216  
QY 241 LQASPPGYLTALRMFNRTYKLSYSYSLGLGMSARLALIGVEGQPAKDGKELVSPCLSP 300  
DB 217 LEQTPRGILTSFEMFNSTYKLTHTSYLGFGLKAARLALIGALETE-GLDGHTRFSACLPR 275  
QY 301 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTBEVKHVDFAFSYYDL 360  
DB 276 WLEAEWTFGGVYKQYQGNQGEVGFPCYAEVLRVGRGLHQPEVQSGSFYAFSYYDR 335  
QY 361 AAGVGLIDAEKGSILVGDFFELAAKYVCRLETOPQSSPFCMDLTVYSLLQF-FGPR 419  
DB 336 AVDTMDIDYEKGSILKVEDFEKRAEVCNDLENFTSGSPFLCMDLSTYITALLKDGFGFAD 395  
QY 420 SKYLKTRKIDNVTSMALGALFYHIDSL 448  
DB 396 STVLQTLTKVNNIETGMALGATFHLLQSL 424

RESULT 15  
US-10-286-926-5  
Sequence 5, Application US/10286926  
Publication No. US20030175752A1  
GENERAL INFORMATION:  
APPLICANT: Ford, John  
APPLICANT: Mulero, Julio  
APPLICANT: Yeung, George  
TITLE OF INVENTION: Methods and Materials Relating to CD39-Like  
Polypeptides

FILE REFERENCE: 28110/36457CON  
CURRENT APPLICATION NUMBER: US/10/286,926  
CURRENT FILING DATE: 2002-11-01  
PRIOR APPLICATION NUMBER: 09/557,800  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/481,238  
PRIOR FILING DATE: 2000-01-11  
PRIOR APPLICATION NUMBER: 09/370,265  
PRIOR FILING DATE: 1999-08-09  
PRIOR APPLICATION NUMBER: PCT/US99/16180  
PRIOR FILING DATE: 1999-07-16  
PRIOR APPLICATION NUMBER: 09/350836  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: 09/273447  
PRIOR FILING DATE: 1999-03-19  
PRIOR APPLICATION NUMBER: 09/122449  
PRIOR FILING DATE: 1998-07-24  
PRIOR APPLICATION NUMBER: 09/244444  
PRIOR FILING DATE: 1999-02-04  
PRIOR APPLICATION NUMBER: 09/118,205  
PRIOR FILING DATE: 1998-07-16  
NUMBER OF SEQ ID NOS: 54  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 5  
LENGTH: 428  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-286-926-5

Query Match 42.3%; Score 999; DB 14; Length 428;  
Best Local Similarity 52.4%; Pred. No. 1.4e-92;  
Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;  
QY 62 PLGTADGHEVFFYGMFAGSTGTGVHVFQFT-RRPREPTLTHETFAVKXGLSAYADD 120  
DB 40 PIVNSA---STLYGIMFAGSTGTGRHIVYTFVQKMPGQLPILGGEVFDVSKGLSAYDQ 96  
QY 121 VEKSAQIRELLDVAKODIPDFFWKATPLVKATAGRLLPGEKAKOKLQKYKEVFKASP 180  
DB 97 PROGAEITVOGLLEVAKDSIPRSHMKKTPLVKATAGRLLPGEKAKALLFEVKEIFPKSP 156  
QY 161 FLVGGDCVSIANGTDEGVSAWITINFLTGLSGKTPGSSVGMULDGGSGTQIAFLPREVGT 240  
DB 157 FLVPGKSVSINDGSDGGLAWTVNFLTQGLHGRQETVGLDGGASTQITFLPOFEKT 216  
QY 241 LQASPPGYLTALRMFNRTYKLSYSYSLGLGMSARLALIGVEGQPAKDGKELVSPCLSP 300  
DB 217 LEQTPRGILTSFEMFNSTYKLTHTSYLGFGLKAARLALIGALETE-GLDGHTRFSACLPR 275  
QY 301 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTBEVKHVDFAFSYYDL 360  
DB 276 WLEAEWTFGGVYKQYQGNQGEVGFPCYAEVLRVGRGLHQPEVQSGSFYAFSYYDR 335  
QY 361 AAGVGLIDAEKGSILVGDFFELAAKYVCRLETOPQSSPFCMDLTVYSLLQF-FGPR 419  
DB 336 AVDTMDIDYEKGSILKVEDFEKRAEVCNDLENFTSGSPFLCMDLSTYITALLKDGFGFAD 395  
QY 420 SKYLKTRKIDNVTSMALGALFYHIDSL 448  
DB 396 STVLQTLTKVNNIETGMALGATFHLLQSL 424

Search completed: June 8, 2004, 09:59:11  
Job time: 51 secs

Tue Jun 8 12:10:38 2004

us-09-905-589a-2.ra1

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OW protein - protein search, using sw model

Run on: June 8, 2004, 09:48:55; Search time 22 Seconds  
(without alignments)  
1070.066 Million cell updates/sec

Title: US-09-905-589A-2

Perfect score: 2364  
Sequence: 1 MKTSHNGSLRVAKVAVPLG.....ALGAIIFYIDSLNRQSPAS 456

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Issued Patents AA:\*

1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/6C.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/6D.COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2364	100.0	456	US-09-240-639-2	Sequence 2, Appli
2	2364	100.0	484	US-09-608-285A-27	Sequence 27, Appli
3	2364	100.0	484	US-09-370-265-27	Sequence 27, Appli
4	2364	100.0	484	US-09-557-800C-27	Sequence 27, Appli
5	2364	100.0	484	US-09-370-625A-27	Sequence 27, Appli
6	2003	84.7	471	US-09-608-285A-60	Sequence 60, Appli
7	999	42.3	428	US-09-608-285A-3	Sequence 3, Appli
8	999	42.3	428	US-09-608-285A-5	Sequence 5, Appli
9	999	42.3	428	US-09-240-639-6	Sequence 6, Appli
10	999	42.3	428	US-09-240-639-9	Sequence 9, Appli
11	999	42.3	428	US-09-350-836B-3	Sequence 3, Appli
12	999	42.3	428	US-09-350-836B-5	Sequence 5, Appli
13	999	42.3	428	US-09-370-265-3	Sequence 3, Appli
14	999	42.3	428	US-09-370-265-5	Sequence 5, Appli
15	999	42.3	428	US-09-557-800C-3	Sequence 3, Appli
16	999	42.3	428	US-09-557-800C-5	Sequence 5, Appli
17	999	42.3	428	US-09-370-625A-3	Sequence 3, Appli
18	999	42.3	428	US-09-370-625A-5	Sequence 5, Appli
19	999	42.1	428	US-09-608-285A-7	Sequence 7, Appli
20	999	42.1	428	US-09-350-836B-7	Sequence 7, Appli
21	999	42.1	428	US-09-370-265-7	Sequence 7, Appli
22	999	42.1	428	US-09-557-800C-7	Sequence 7, Appli
23	999	42.1	428	US-09-370-625A-7	Sequence 7, Appli
24	909	38.5	405	US-09-608-285A-25	Sequence 25, Appli
25	909	38.5	405	US-09-370-265-25	Sequence 25, Appli
26	909	38.5	405	US-09-557-800C-25	Sequence 25, Appli
27	909	38.5	405	US-09-370-625A-25	Sequence 25, Appli

28	909	38.5	465	US-09-557-800C-56	Sequence 56, Appli
29	909	38.5	465	US-09-370-625A-39	Sequence 39, Appli
30	904	38.2	465	US-09-240-639-8	Sequence 8, Appli
31	525.5	22.2	467	US-09-129-112-19	Sequence 19, Appli
32	506	21.4	462	US-09-129-112-2	Sequence 2, Appli
33	501	21.2	459	US-09-129-112-9	Sequence 9, Appli
34	485	20.5	454	US-09-240-639-11	Sequence 11, Appli
35	466	19.7	473	US-09-240-639-12	Sequence 12, Appli
36	463	19.6	462	US-09-129-112-15	Sequence 15, Appli
37	442.5	18.7	455	US-09-240-639-10	Sequence 10, Appli
38	314.5	13.3	529	US-09-240-639-4	Sequence 4, Appli
39	285.5	12.1	502	US-09-557-800C-55	Sequence 55, Appli
40	285.5	12.1	502	US-09-370-625A-38	Sequence 38, Appli
41	285.5	12.1	510	US-08-930-921-1	Sequence 1, Appli
42	216.5	9.2	148	US-09-240-639-17	Sequence 17, Appli
43	184	7.8	150	US-09-240-639-16	Sequence 16, Appli
44	183.5	7.8	154	US-09-240-639-14	Sequence 14, Appli
45	175.5	7.4	153	US-09-240-639-15	Sequence 15, Appli

## ALIGNMENTS

RESULT 1  
US-09-240-639-2  
Sequence 2, Application US/09240639  
Patent No. 635047  
GENERAL INFORMATION:  
APPLICANT: Chadwick, Brian Paul  
APPLICANT: Frischau, Anna-Maria  
TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO CD39-LIKE  
TITLE OF INVENTION: POLYPEPTIDES AND NUCLEIC ACIDS  
FILE REFERENCE: 9598-066  
CURRENT APPLICATION NUMBER: US/09/240, 639  
CURRENT FILING DATE: 1998-01-29  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 2  
LENGTH: 456  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-240-639-2

Query Match 100.0%; Score 2364; DB 4; Length 456;  
Best Local Similarity 100.0%; Pred. No. 5.8e-240;  
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MKTSHNGSLRVAKVAVPLG	LCVGVFIYVYIKMHRATQAFPSITRAFGARWQQA	60
DB	1	MKTSHNGSLRVAKVAVPLG	LCVGVFIYVYIKMHRATQAFPSITRAFGARWQQA	60
QY	61	SPGTADGHEVYVYIMFAGSTGT	RVHFOFTRPREPTLTHETFKAVKGLSAVDD	120
DB	61	SPGTADGHEVYVYIMFAGSTGT	RVHFOFTRPREPTLTHETFKAVKGLSAVDD	120
QY	121	VERSAQIGHEILDVAQDIP	DFWKATPLVLAATGRLLPBKAOKLQKVEYFKASP	180
DB	121	VERSAQIGHEILDVAQDIP	DFWKATPLVLAATGRLLPBKAOKLQKVEYFKASP	180
QY	181	FLVGGDCVSTIMNGTDSGVSAM	TINVLNGSLKTPGSSVGMDLGGSGTQIAFLPVEGT	240
DB	181	FLVGGDCVSTIMNGTDSGVSAM	TINVLNGSLKTPGSSVGMDLGGSGTQIAFLPVEGT	240
QY	241	LQASPPGYLTALMFRTYKLS	YSYSLGLGMSARLAIIGVGGQPAKDGKELVSPCLSP	300
DB	241	LQASPPGYLTALMFRTYKLS	YSYSLGLGMSARLAIIGVGGQPAKDGKELVSPCLSP	300
QY	301	SFGGEVHEVYRVSGQRAASL	HELCAARSEVQNVRHTEEVKADVFPAFSYYDL	360
DB	301	SFGGEVHEVYRVSGQRAASL	HELCAARSEVQNVRHTEEVKADVFPAFSYYDL	360
QY	361	AAGVGILDEKGGSLVVGDFE	IAKTYVCTLETOPSSPSCMDLTVYSLIQERFPRS	420
DB	361	AAGVGILDEKGGSLVVGDFE	IAKTYVCTLETOPSSPSCMDLTVYSLIQERFPRS	420

Db 361 AAGVGLIDAEKGGSLVVGDFEIAAKVCRTELETPQSSPFCMDLTVVSLLLQBFQFPRS 420  
Qy 421 KVLKTRKIDNVETSWALGAIFHYIDSLNRQKSPAS 456  
Db 421 KVLKTRKIDNVETSWALGAIFHYIDSLNRQKSPAS 456  
RESULT 2  
US-09-608-285A-27  
; Sequence 27, Application US/09608285A  
; Patent No. 6335013  
; GENERAL INFORMATION:  
; APPLICANT: Ford, John  
; APPLICANT: Muleto, George  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO CD39-LIKE  
; FILE REFERENCE: 28110/36570  
; CURRENT APPLICATION NUMBER: US/09/608,285A  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: 09/583,231  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 09/557,800  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/481,238  
; PRIOR FILING DATE: 2000-01-11  
; PRIOR APPLICATION NUMBER: 09/370,265  
; PRIOR FILING DATE: 1999-08-09  
; PRIOR APPLICATION NUMBER: PCT/US99/16180  
; PRIOR FILING DATE: 1999-07-16  
; PRIOR APPLICATION NUMBER: 09/350,836  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: 09/273,447  
; PRIOR FILING DATE: 1999-03-19  
; PRIOR APPLICATION NUMBER: 09/244,444  
; PRIOR FILING DATE: 1999-02-04  
; PRIOR APPLICATION NUMBER: 09/122,449  
; PRIOR FILING DATE: 1998-07-24  
; PRIOR APPLICATION NUMBER: 09/118,205  
; NUMBER OF SEQ ID NOS: 60  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 27  
; LENGTH: 484  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-608-285A-27  
Query Match 100.0%; Score 2364; DB 4; Length 484;  
Best Local Similarity 100.0%; Pred. No. 6.4e-240;  
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MRKISNHSGLRVAKVAYPLGLCVGVFIYVAYIKWHRATATQAFPSITRAAPGARWQQA 60  
Db 29 MRKISNHSGLRVAKVAYPLGLCVGVFIYVAYIKWHRATATQAFPSITRAAPGARWQQA 88  
Qy 61 SPLGTAADGHEVYGINFMDAGSTGTRVHVQFTRPPRETPTLTHETFKAVKPLSAYADD 120  
Db 89 SPLGTAADGHEVYGINFMDAGSTGTRVHVQFTRPPRETPTLTHETFKAVKPLSAYADD 148  
Qy 121 VEKSAQGIREDLDVAKODIPDFWKATPLVKATAGLRLLPGEKAQKLLQKVEVFKASP 180  
Db 149 VEKSAQGIREDLDVAKODIPDFWKATPLVKATAGLRLLPGEKAQKLLQKVEVFKASP 208  
Qy 181 FLVGDDCVSINMGTDGVSAMITINFLTGLKTPGGSSVGMGLDLGGGSTQIAFLPRVEGT 240  
Db 209 FLVGDDCVSINMGTDGVSAMITINFLTGLKTPGGSSVGMGLDLGGGSTQIAFLPRVEGT 268  
Qy 241 LOASPPGYLTALRMFNRTYKLYSYSLGLGMSARLAILGGVGEQPAKDGKELVSPCLSP 300  
Db 269 FLVGDDCVSINMGTDGVSAMITINFLTGLKTPGGSSVGMGLDLGGGSTQIAFLPRVEGT 328  
Qy 301 LQASPPGYLTALRMFNRTYKLYSYSLGLGMSARLAILGGVGEQPAKDGKELVSPCLSP 300  
Db 269 LQASPPGYLTALRMFNRTYKLYSYSLGLGMSARLAILGGVGEQPAKDGKELVSPCLSP 328  
Qy 301 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTVEEVKHVDFYAFSYYDL 360

Db 329 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTVEEVKHVDFYAFSYYDL 388  
Qy 361 AAGVGLIDAEKGGSLVVGDFEIAAKVCRTELETPQSSPFCMDLTVVSLLLQBFQFPRS 420  
Db 389 AAGVGLIDAEKGGSLVVGDFEIAAKVCRTELETPQSSPFCMDLTVVSLLLQBFQFPRS 448  
Qy 421 KVLKTRKIDNVETSWALGAIFHYIDSLNRQKSPAS 456  
Db 449 KVLKTRKIDNVETSWALGAIFHYIDSLNRQKSPAS 484  
RESULT 3  
US-09-370-265-27  
; Sequence 27, Application US/09370265  
; Patent No. 6447771  
; GENERAL INFORMATION:  
; APPLICANT: Ford, John  
; APPLICANT: Muleto, Julio  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE  
; FILE REFERENCE: 28111/35908  
; CURRENT APPLICATION NUMBER: US/09/370,265  
; CURRENT FILING DATE: 1999-08-09  
; EARLIER APPLICATION NUMBER: PCT/US99/16180  
; EARLIER FILING DATE: 1999-07-16  
; EARLIER APPLICATION NUMBER: 09/350,836  
; EARLIER FILING DATE: 1999-07-09  
; EARLIER APPLICATION NUMBER: 09/273,447  
; EARLIER FILING DATE: 1999-03-19  
; EARLIER APPLICATION NUMBER: 09/244,444  
; EARLIER FILING DATE: 1999-02-04  
; EARLIER APPLICATION NUMBER: 09/122,449  
; EARLIER FILING DATE: 1998-07-24  
; EARLIER APPLICATION NUMBER: 09/118,205  
; EARLIER FILING DATE: 1998-07-16  
; NUMBER OF SEQ ID NOS: 37  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 27  
; LENGTH: 484  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-370-265-27  
Query Match 100.0%; Score 2364; DB 4; Length 484;  
Best Local Similarity 100.0%; Pred. No. 6.4e-240;  
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MRKISNHSGLRVAKVAYPLGLCVGVFIYVAYIKWHRATATQAFPSITRAAPGARWQQA 60  
Db 29 MRKISNHSGLRVAKVAYPLGLCVGVFIYVAYIKWHRATATQAFPSITRAAPGARWQQA 88  
Qy 61 SPLGTAADGHEVYGINFMDAGSTGTRVHVQFTRPPRETPTLTHETFKAVKPLSAYADD 120  
Db 89 SPLGTAADGHEVYGINFMDAGSTGTRVHVQFTRPPRETPTLTHETFKAVKPLSAYADD 148  
Qy 121 VEKSAQGIREDLDVAKODIPDFWKATPLVKATAGLRLLPGEKAQKLLQKVEVFKASP 180  
Db 149 VEKSAQGIREDLDVAKODIPDFWKATPLVKATAGLRLLPGEKAQKLLQKVEVFKASP 208  
Qy 181 FLVGDDCVSINMGTDGVSAMITINFLTGLKTPGGSSVGMGLDLGGGSTQIAFLPRVEGT 240  
Db 209 FLVGDDCVSINMGTDGVSAMITINFLTGLKTPGGSSVGMGLDLGGGSTQIAFLPRVEGT 268  
Qy 241 LOASPPGYLTALRMFNRTYKLYSYSLGLGMSARLAILGGVGEQPAKDGKELVSPCLSP 300  
Db 269 LOASPPGYLTALRMFNRTYKLYSYSLGLGMSARLAILGGVGEQPAKDGKELVSPCLSP 328  
Qy 301 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTVEEVKHVDFYAFSYYDL 360  
Db 329 SFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTVEEVKHVDFYAFSYYDL 388  
Qy 361 AAGVGLIDAEKGGSLVVGDFEIAAKVCRTELETPQSSPFCMDLTVVSLLLQBFQFPRS 420

Db 389 AAGVGLIDAEKGSLLVGGDFEIAKYVCRLETPQSSPSCMDLTVYSLLOEFGPPRS 448  
Qy 421 KVLKLTTRKIDNVTETSWALGAIFFHYIDSLNRQKSPAS 456  
Db 449 KVLKLTTRKIDNVTETSWALGAIFFHYIDSLNRQKSPAS 484

RESULT 4  
US-09-557-800C-27  
Sequence 27, Application US/09557800C  
Patent No. 6476211  
GENERAL INFORMATION:  
APPLICANT: Ford, John  
APPLICANT: Mulero, Julio  
APPLICANT: Yeung, George  
TITLE OF INVENTION: Methods and Materials Relating to CD39-Like  
FILE REFERENCE: 28110/36457  
CURRENT APPLICATION NUMBER: US/09/557,800C  
CURRENT FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/461,238  
PRIOR FILING DATE: 2000-01-11  
PRIOR APPLICATION NUMBER: 09/370,265  
PRIOR FILING DATE: 1999-08-09  
PRIOR APPLICATION NUMBER: PCT/US99/16180  
PRIOR FILING DATE: 1999-07-16  
PRIOR APPLICATION NUMBER: 09/350836  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: 09/273447  
PRIOR FILING DATE: 1999-03-19  
PRIOR APPLICATION NUMBER: 09/122449  
PRIOR FILING DATE: 1998-07-24  
PRIOR APPLICATION NUMBER: 09/244444  
PRIOR FILING DATE: 1999-02-04  
PRIOR APPLICATION NUMBER: 09/118,205  
PRIOR FILING DATE: 1998-07-16  
NUMBER OF SEQ ID NOS: 56  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 27  
LENGTH: 484  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-557-800C-27

Query Match 100.0%; Score 2364; DB 4; Length 484;  
Best Local Similarity 100.0%; Pred. No. 6, 4e-240;  
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MRKISNHSLSLVAKAVPLGLCVGFVTVAYIKMHRATATQAFPSITRAAPGARWGOAH 60  
Db 29 MRKISNHSLSLVAKAVPLGLCVGFVTVAYIKMHRATATQAFPSITRAAPGARWGOAH 88  
Qy 61 SPLGTADGHEVFGIMFDAGSTGRVAVFOFTPPREPTTLTHETFRKAVPGLSAYAD 120  
Db 89 SPLGTADGHEVFGIMFDAGSTGRVAVFOFTPPREPTTLTHETFRKAVPGLSAYAD 148  
Qy 121 VEKSAQGIREDLDVAKODIPDFWKATPLVLTAKATAGRLLPGEKAOKLLQKVEVFAS 180  
Db 149 VEKSAQGIREDLDVAKODIPDFWKATPLVLTAKATAGRLLPGEKAOKLLQKVEVFAS 208  
Qy 181 FLVGGDCVSIKNGDEGVASAMITINFLTGSIKTPGSSVGMULDGGSTQIAFLPVEGT 240  
Db 209 FLVGGDCVSIKNGDEGVASAMITINFLTGSIKTPGSSVGMULDGGSTQIAFLPVEGT 268  
Qy 241 LQASPPGYLTALRMFNRTYKLYSYSLGLGMSARLAILGVEGQPAKDGKEIVSPCLSP 300  
Db 269 LQASPPGYLTALRMFNRTYKLYSYSLGLGMSARLAILGVEGQPAKDGKEIVSPCLSP 328  
Qy 301 SFKGEWEHAEVTVYVSGQKAAASLHELCAARVSEVLTQNRVHRTTEVXVDFYAFSYYDL 360  
Db 329 SFKGEWEHAEVTVYVSGQKAAASLHELCAARVSEVLTQNRVHRTTEVXVDFYAFSYYDL 388

Qy 361 AAGVGLIDAEKGSLLVGGDFEIAKYVCRLETPQSSPSCMDLTVYSLLOEFGPPRS 420  
Db 389 AAGVGLIDAEKGSLLVGGDFEIAKYVCRLETPQSSPSCMDLTVYSLLOEFGPPRS 448  
Qy 421 KVLKLTTRKIDNVTETSWALGAIFFHYIDSLNRQKSPAS 456  
Db 449 KVLKLTTRKIDNVTETSWALGAIFFHYIDSLNRQKSPAS 484

RESULT 5  
US-09-370-625A-27  
Sequence 27, Application US/09370625A  
Patent No. 660032  
GENERAL INFORMATION:  
APPLICANT: Ford, John  
APPLICANT: Mulero, Julio  
APPLICANT: Yeung, George  
TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE POLYPEPTIDES  
FILE REFERENCE: 28110/35908  
CURRENT APPLICATION NUMBER: US/09/370,625A  
CURRENT FILING DATE: 1999-08-09  
PRIOR APPLICATION NUMBER: PCT/US99/16180  
PRIOR FILING DATE: 1999-07-16  
PRIOR APPLICATION NUMBER: 09/350,836  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: 09/273,447  
PRIOR FILING DATE: 1999-03-19  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 27  
LENGTH: 484  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-370-625A-27

Query Match 100.0%; Score 2364; DB 4; Length 484;  
Best Local Similarity 100.0%; Pred. No. 6, 4e-240;  
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRKISNHSLSLVAKAVPLGLCVGFVTVAYIKMHRATATQAFPSITRAAPGARWGOAH 60  
Db 29 MRKISNHSLSLVAKAVPLGLCVGFVTVAYIKMHRATATQAFPSITRAAPGARWGOAH 88  
Qy 61 SPLGTADGHEVFGIMFDAGSTGRVAVFOFTPPREPTTLTHETFRKAVPGLSAYAD 120  
Db 89 SPLGTADGHEVFGIMFDAGSTGRVAVFOFTPPREPTTLTHETFRKAVPGLSAYAD 148  
Qy 121 VEKSAQGIREDLDVAKODIPDFWKATPLVLTAKATAGRLLPGEKAOKLLQKVEVFAS 180  
Db 149 VEKSAQGIREDLDVAKODIPDFWKATPLVLTAKATAGRLLPGEKAOKLLQKVEVFAS 208  
Qy 181 FLVGGDCVSIKNGDEGVASAMITINFLTGSIKTPGSSVGMULDGGSTQIAFLPVEGT 240  
Db 209 FLVGGDCVSIKNGDEGVASAMITINFLTGSIKTPGSSVGMULDGGSTQIAFLPVEGT 268  
Qy 241 LQASPPGYLTALRMFNRTYKLYSYSLGLGMSARLAILGVEGQPAKDGKEIVSPCLSP 300  
Db 269 LQASPPGYLTALRMFNRTYKLYSYSLGLGMSARLAILGVEGQPAKDGKEIVSPCLSP 328  
Qy 301 SFKGEWEHAEVTVYVSGQKAAASLHELCAARVSEVLTQNRVHRTTEVXVDFYAFSYYDL 360  
Db 329 SFKGEWEHAEVTVYVSGQKAAASLHELCAARVSEVLTQNRVHRTTEVXVDFYAFSYYDL 388  
Qy 361 AAGVGLIDAEKGSLLVGGDFEIAKYVCRLETPQSSPSCMDLTVYSLLOEFGPPRS 420  
Db 389 AAGVGLIDAEKGSLLVGGDFEIAKYVCRLETPQSSPSCMDLTVYSLLOEFGPPRS 448  
Qy 421 KVLKLTTRKIDNVTETSWALGAIFFHYIDSLNRQKSPAS 456  
Db 449 KVLKLTTRKIDNVTETSWALGAIFFHYIDSLNRQKSPAS 484

RESULT 6  
US-09-608-285A-60



; Sequence 60, Application US/09608285A  
; Patent No. 6335013  
; GENERAL INFORMATION:  
; APPLICANT: Ford, John  
; APPLICANT: Mulero, Julio  
; APPLICANT: Yeung, George  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO CD39-LIKE  
; TITLE OF INVENTION: POLYPEPTIDES  
; FILE REFERENCE: 28110/36570  
; CURRENT APPLICATION NUMBER: US/09/608,285A  
; CURRENT FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: 09/583,231  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 09/557,800  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/481,238  
; PRIOR FILING DATE: 2000-01-11  
; PRIOR APPLICATION NUMBER: 09/370,265  
; PRIOR FILING DATE: 1999-08-09  
; PRIOR APPLICATION NUMBER: PCT/US99/16180  
; PRIOR FILING DATE: 1999-07-16  
; PRIOR APPLICATION NUMBER: 09/350,836  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: 09/273,447  
; PRIOR FILING DATE: 1999-03-19  
; PRIOR APPLICATION NUMBER: 09/244,444  
; PRIOR FILING DATE: 1999-02-04  
; PRIOR APPLICATION NUMBER: 09/122,449  
; PRIOR FILING DATE: 1998-07-24  
; PRIOR APPLICATION NUMBER: 09/118,205  
; PRIOR FILING DATE: 1998-07-16  
; NUMBER OF SEQ ID NOS: 60  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 60  
; LENGTH: 471  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-608-285A-60

Query Match 84.7%; Score 2003; DB 4; Length 471;  
Best Local Similarity 98.5%; Pred. No. 6.3e-202;  
Matches 388; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 MRKISNHGSLRVAKVAYPLGCVGVFIYVAYIKWHRATATQAFPSITRAAPGARWQQAH 60  
Db 29 MRKISNHGSLRVAKVAYPLGCVGVFIYVAYIKWHRATATQAFPSITRAAPGARWQQAH 88  
Qy 61 SPLGTAADGHEVFGIMFDAGSTGTRVHVFOFTPRPREPTTLTHTFKAVKPLGSLAYADD 120  
Db 89 SPLGTAADGHEVFGIMFDAGSTGTRVHVFOFTPRPREPTTLTHTFKAVKPLGSLAYADD 148  
Qy 121 VEKSAQGIREDLLVAKQDIPFDWKATPLVLKATAGLRLLPGEKAKQLKQKVEVFKASP 180  
Db 149 VEKSAQGIREDLLVAKQDIPFDWKATPLVLKATAGLRLLPGEKAKQLKQKVEVFKASP 208  
Qy 181 FLVGDDCVSIMNGTDEGVSAWITINFLTGLSKTPGSSVGMGLDGGSTQIAFLPRVEGT 240  
Db 209 FLVGDDCVSIMNGTDEGVSAWITINFLTGLSKTPGSSVGMGLDGGSTQIAFLPRVEGT 268  
Qy 241 LQASPPGYLTALRMFNRTYKLYSYSLGLGMSARLAILGGVGEQPAKDGKELVSPCLSP 300  
Db 269 LQASPPGYLTALRMFNRTYKLYSYSLGLGMSARLAILGGVGEQPAKDGKELVSPCLSP 328  
Qy 301 SFKGEWEHAEVTVRVSGKAAASLHELCAARVSEVLQNRVHRTTEVKHVDYFAFSYYIDL 360  
Db 329 SFKGEWEHAEVTVRVSGKAAASLHELCAARVSEVLQNRVHRTTEVKHVDYFAFSYYIDL 388  
Qy 361 AAGVGLIDAEKGGSLVVGDFEIAAKVCRTELTQ 394  
Db 389 AAGVGLIDAEKGGSLVVGDFEIAAKVCRTELTQ 422

RESULT 7

US-09-608-285A-3  
; Sequence 3, Application US/09608285A  
; Patent No. 6335013  
; GENERAL INFORMATION:  
; APPLICANT: Ford, John  
; APPLICANT: Mulero, Julio  
; APPLICANT: Yeung, George  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO CD39-LIKE  
; TITLE OF INVENTION: POLYPEPTIDES  
; FILE REFERENCE: 28110/36570  
; CURRENT APPLICATION NUMBER: US/09/608,285A  
; CURRENT FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: 09/583,231  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 09/557,800  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/481,238  
; PRIOR FILING DATE: 2000-01-11  
; PRIOR APPLICATION NUMBER: 09/370,265  
; PRIOR FILING DATE: 1999-08-09  
; PRIOR APPLICATION NUMBER: PCT/US99/16180  
; PRIOR FILING DATE: 1999-07-16  
; PRIOR APPLICATION NUMBER: 09/350,836  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: 09/273,447  
; PRIOR FILING DATE: 1999-03-19  
; PRIOR APPLICATION NUMBER: 09/244,444  
; PRIOR FILING DATE: 1999-02-04  
; PRIOR APPLICATION NUMBER: 09/122,449  
; PRIOR FILING DATE: 1998-07-24  
; PRIOR APPLICATION NUMBER: 09/118,205  
; PRIOR FILING DATE: 1998-07-16  
; NUMBER OF SEQ ID NOS: 60  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 428  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-608-285A-3

Query Match 42.3%; Score 999; DB 4; Length 428;  
Best Local Similarity 52.4%; Pred. No. 2.8e-96;  
Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;  
Qy 62 PLGTAADGHEVFGIMFDAGSTGTRVHVFOFTPRPREPTTLTHTFKAVKPLGSLAYADD 120  
Db 40 PINVSA--STLYGIMFDAGSTGTRVHVTFVQKMPGQLPILGEVFDVSKPGLSAFVDQ 96  
Qy 121 VEKSAQGIREDLLVAKQDIPFDWKATPLVLKATAGLRLLPGEKAKQLKQKVEVFKASP 180  
Db 97 PRQAEIVQGLLELVAKQDIPRSHWKTFFVVLKATAGLRLLPGEKAKQLKQKVEVFKASP 156  
Qy 181 FLVGDDCVSIMNGTDEGVSAWITINFLTGLSKTPGSSVGMGLDGGSTQIAFLPRVEGT 240  
Db 157 FLVPGKSVSIMDGSDEGILAWTVNFLTGLHGHROBTGTVGLDGGASTQITLPPQPEKT 216  
Qy 241 LQASPPGYLTALRMFNRTYKLYSYSLGLGMSARLAILGGVGEQPAKDGKELVSPCLSP 300  
Db 217 LEQTFPGYLTALRMFNRTYKLYSYSLGLGMSARLAILGGVGEQPAKDGKELVSPCLSP 275  
Qy 301 SFKGEWEHAEVTVRVSGKAAASLHELCAARVSEVLQNRVHRTTEVKHVDYFAFSYYIDL 360  
Db 276 WLEAEWIFGGYKQYGGQSGEVGFCPCYAEVLVRVVRGKLHQPEEVQSGFYAFSYYIDR 335  
Qy 361 AAGVGLIDAEKGGSLVVGDFEIAAKVCRTELTQPOQSSPSCMDLTVVSLLLQE-FCGPR 419  
Db 336 AVDTMDIDYKGGILKVEDFERKAREVCDNLENTSGSPFLCMDLSYITALLKDGPGFAD 395  
Qy 420 SKVLKTRKIDNVETSWALGAIFHYIDSL 448  
Db 396 STVLQILTKVNNIETGHALGATFHLQSL 424



```
Query Match 42.3%; Score 999; DB 4; Length 428;
Best Local Similarity 52.4%; Pred. No. 2.8e-96;
Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;

QY 62 PLGTAADGHEVYFYGIMFDAGSTGRVHVFOFT-RPPRETPLTHETFKAVKPGLSAYADD 120
DB 40 PINVSA---STLYGIMFDAGSTGTRIHVYTFVQKMPGQLPILGEVFDVSKVPGLSAFVDQ 96

QY 121 VEKSAQGIREDLLDYAKODIPDFWKATPLVLKATAGLRLLPGEKAOKLQKVEFKASP 180
DB 97 PKQGAETVQGLLEVAKODISPRSHWKKTPVVLKATAGLRLLPGEKAOKLQKVEFKASP 156

QY 181 FLVGDDCVSINMGTDGVSAMITINFLTGLSGKTPGSSVGMGLDGGSGTOIAFLPRVEGT 240
DB 157 FLVPGKSVSIMDGSDEGILAWTVNFLTGLQHGHRQETVGTDLGGASTQITFLPQPEKT 216

QY 241 LQASPPGYLTALRMFNRTYKLYSYLGLGMSARLAILGSGVEGQPAKDGKELVSPCLSP 300
DB 217 LEQTPRGYLTSEFMFNSTYKLYTHSYLGFGLKAARLATLGALETE-GTDGHTFSAACLP 275

QY 301 SFKGEWEHAEVTVYRVSGQKAAASLHELCAARVSEVLQNRVHRTVEVKHVDVFAFSYYDL 360
DB 276 WLEAEWIFGGVKYQYGGNQEVEGFPCYAEVLVRVGRKLHQPEEVQSGSFYAFSYYDR 335

QY 361 AAGVGLIDAEKGGSLVGDDEIAAKYVCRILETOQSPSPSCMDLTYVSLLOE-FGPPR 419
DB 336 AVDTMDIDYEGKGLKVEDFERKAREVCNLENFTSGSPFLCMLDLYTALLKDGFGFAD 395

QY 420 SKVLKLRKIDNVTSWALGAIPHVIDSL 448
DB 396 STVLQLTCKVNNIETGALGATFHLQSL 424

RESULT 11
US-09-350-836B-3
; Sequence 3, Application US/09350836B
; Patent No. 6387645
; GENERAL INFORMATION:
; APPLICANT: Mulero, Julio
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE
; FILE REFERENCE: 28110/35761
; CURRENT APPLICATION NUMBER: US/09/350,836B
; CURRENT FILING DATE: 1999-07-09
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 09/118,205
; PRIOR FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 09/122,449
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 09/244,444
; PRIOR FILING DATE: 1999-02-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 428
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-350-836B-3

Query Match 42.3%; Score 999; DB 4; Length 428;
Best Local Similarity 52.4%; Pred. No. 2.8e-96;
Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;

QY 62 PLGTAADGHEVYFYGIMFDAGSTGRVHVFOFT-RPPRETPLTHETFKAVKPGLSAYADD 120
DB 40 PINVSA---STLYGIMFDAGSTGTRIHVYTFVQKMPGQLPILGEVFDVSKVPGLSAFVDQ 96

QY 121 VEKSAQGIREDLLDYAKODIPDFWKATPLVLKATAGLRLLPGEKAOKLQKVEFKASP 180
DB 97 PKQGAETVQGLLEVAKODISPRSHWKKTPVVLKATAGLRLLPGEKAOKLQKVEFKASP 156

QY 181 FLVGDDCVSINMGTDGVSAMITINFLTGLSGKTPGSSVGMGLDGGSGTOIAFLPRVEGT 240
DB 157 FLVPGKSVSIMDGSDEGILAWTVNFLTGLQHGHRQETVGTDLGGASTQITFLPQPEKT 216

QY 241 LQASPPGYLTALRMFNRTYKLYSYLGLGMSARLAILGSGVEGQPAKDGKELVSPCLSP 300
DB 217 LEQTPRGYLTSEFMFNSTYKLYTHSYLGFGLKAARLATLGALETE-GTDGHTFSAACLP 275

QY 301 SFKGEWEHAEVTVYRVSGQKAAASLHELCAARVSEVLQNRVHRTVEVKHVDVFAFSYYDL 360
DB 276 WLEAEWIFGGVKYQYGGNQEVEGFPCYAEVLVRVGRKLHQPEEVQSGSFYAFSYYDR 335

QY 361 AAGVGLIDAEKGGSLVGDDEIAAKYVCRILETOQSPSPSCMDLTYVSLLOE-FGPPR 419
DB 336 AVDTMDIDYEGKGLKVEDFERKAREVCNLENFTSGSPFLCMLDLYTALLKDGFGFAD 395

QY 420 SKVLKLRKIDNVTSWALGAIPHVIDSL 448
DB 396 STVLQLTCKVNNIETGALGATFHLQSL 424
```

```
QY 181 FLVGDDCVSINMGTDGVSAMITINFLTGLSGKTPGSSVGMGLDGGSGTOIAFLPRVEGT 240
DB 157 FLVPGKSVSIMDGSDEGILAWTVNFLTGLQHGHRQETVGTDLGGASTQITFLPQPEKT 216

QY 241 LQASPPGYLTALRMFNRTYKLYSYLGLGMSARLAILGSGVEGQPAKDGKELVSPCLSP 300
DB 217 LEQTPRGYLTSEFMFNSTYKLYTHSYLGFGLKAARLATLGALETE-GTDGHTFSAACLP 275

QY 301 SFKGEWEHAEVTVYRVSGQKAAASLHELCAARVSEVLQNRVHRTVEVKHVDVFAFSYYDL 360
DB 276 WLEAEWIFGGVKYQYGGNQEVEGFPCYAEVLVRVGRKLHQPEEVQSGSFYAFSYYDR 335

QY 361 AAGVGLIDAEKGGSLVGDDEIAAKYVCRILETOQSPSPSCMDLTYVSLLOE-FGPPR 419
DB 336 AVDTMDIDYEGKGLKVEDFERKAREVCNLENFTSGSPFLCMLDLYTALLKDGFGFAD 395

QY 420 SKVLKLRKIDNVTSWALGAIPHVIDSL 448
DB 396 STVLQLTCKVNNIETGALGATFHLQSL 424

RESULT 12
US-09-350-836B-5
; Sequence 5, Application US/09350836B
; Patent No. 6387645
; GENERAL INFORMATION:
; APPLICANT: Mulero, Julio
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE
; FILE REFERENCE: 28110/35761
; CURRENT APPLICATION NUMBER: US/09/350,836B
; CURRENT FILING DATE: 1999-07-09
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 09/118,205
; PRIOR FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 09/122,449
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 09/244,444
; PRIOR FILING DATE: 1999-02-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 428
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-350-836B-5

Query Match 42.3%; Score 999; DB 4; Length 428;
Best Local Similarity 52.4%; Pred. No. 2.8e-96;
Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;

QY 62 PLGTAADGHEVYFYGIMFDAGSTGRVHVFOFT-RPPRETPLTHETFKAVKPGLSAYADD 120
DB 40 PINVSA---STLYGIMFDAGSTGTRIHVYTFVQKMPGQLPILGEVFDVSKVPGLSAFVDQ 96

QY 121 VEKSAQGIREDLLDYAKODIPDFWKATPLVLKATAGLRLLPGEKAOKLQKVEFKASP 180
DB 97 PKQGAETVQGLLEVAKODISPRSHWKKTPVVLKATAGLRLLPGEKAOKLQKVEFKASP 156

QY 181 FLVGDDCVSINMGTDGVSAMITINFLTGLSGKTPGSSVGMGLDGGSGTOIAFLPRVEGT 240
DB 157 FLVPGKSVSIMDGSDEGILAWTVNFLTGLQHGHRQETVGTDLGGASTQITFLPQPEKT 216

QY 241 LQASPPGYLTALRMFNRTYKLYSYLGLGMSARLAILGSGVEGQPAKDGKELVSPCLSP 300
DB 217 LEQTPRGYLTSEFMFNSTYKLYTHSYLGFGLKAARLATLGALETE-GTDGHTFSAACLP 275

QY 301 SFKGEWEHAEVTVYRVSGQKAAASLHELCAARVSEVLQNRVHRTVEVKHVDVFAFSYYDL 360
DB 276 WLEAEWIFGGVKYQYGGNQEVEGFPCYAEVLVRVGRKLHQPEEVQSGSFYAFSYYDR 335
```

QY 361 AAGVGLIDAEKGGSLVVGDFEIAKVCRTLETQPOSSPCMDLTYVSLLOE-FGPPR 419  
DB 336 AVDTMDIDYKGGILKVEDFERKAREVCNDLENFTSGSPFLCMDLSTYITALLKDGFGFAD 395  
QY 420 SKVLKTRKIDNVTSMALGAI FHYIDSL 448  
DB 396 STVLQTLTKVNNIETGMALGATFHLQSL 424

## RESULT 13

US-09-370-265-3  
; Sequence 3, Application US/09370265  
; Patent No. 6447771  
; GENERAL INFORMATION:  
; APPLICANT: Mulero, John  
; APPLICANT: Mulero, Jullio  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE  
; TITLE OF INVENTION: POLYPEPTIDES  
; FILE REFERENCE: 2811/35908  
; CURRENT APPLICATION NUMBER: US/09/370,265  
; EARLIER FILING DATE: 1999-08-09  
; EARLIER APPLICATION NUMBER: PCT/US99/16180  
; EARLIER FILING DATE: 1999-07-16  
; EARLIER APPLICATION NUMBER: 09/350,836  
; EARLIER FILING DATE: 1999-07-09  
; EARLIER APPLICATION NUMBER: 09/273,447  
; EARLIER FILING DATE: 1999-03-19  
; EARLIER APPLICATION NUMBER: 09/244,444  
; EARLIER FILING DATE: 1999-02-04  
; EARLIER APPLICATION NUMBER: 09/122,449  
; EARLIER FILING DATE: 1998-07-24  
; EARLIER APPLICATION NUMBER: 09/118,205  
; EARLIER FILING DATE: 1998-07-16  
; NUMBER OF SEQ ID NOS: 37  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 428  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-370-265-3

Query Match 42.3%; Score 999; DB 4; Length 428;

Best Local Similarity 52.4%; Pred. No. 2.8e-96;  
Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;

QY 62 PLGTAAADGHEVYFVIMPDAGSTGTRVHVFOFT-RPPRETPLTHETFKAVKPGLSAYVD 120  
DB 40 PINVSA---STLYGIMFDPAGSTGTRIHVYTFVQKMPQGLPLBGEVFDSPKPSAFVDQ 96  
QY 121 VEKSAQGIREFLIDVAKODIPDFWKATPVLKATAGRLLPGEKAKQLLOKYKVEFKASP 180  
DB 97 PKQAEVYQGLLEVAKDISPRSHMKKTPVVKATAGRLLPENKAKALLFEVKEIFRXPSP 156  
QY 181 FLVGGDDCVSINMGDEGVSANITINFLTGSUKTGSSVGMULDGGSGSTQIAFLPRVEGT 240  
DB 157 FLVPRGVSIVMGDSDEGLAVTVNFTLGQJHGRQETVGLDGGASTQITFLPQFEKT 216  
QY 241 LQASPPGYLTALRMFNRTYKLYSYSLGLGMSARLAILGGVEGQPADGKELVSPCLSP 300  
DB 217 LEQTPRGVLTSPFEMFNSTYKLYTHSYLFGKAKARLITGLALETE-GTDGHTFRSACLPR 275  
QY 301 SFKGEWEHAETVYRVSQKAAASLHELCAARYSEVLONRVHRTVEYKVDVPAFSYYIDL 360  
DB 276 WLEAEWIFGVYKYQYGVNOGEVGEPEPCYAEVLRVVRKQLQBPBEVORGSEFYAFSYYYDR 335  
QY 361 AAGVGLIDAEKGGSLVVGDFEIAKVCRTLETQPOSSPCMDLTYVSLLOE-FGPPR 419  
DB 336 AVDTMDIDYKGGILKVEDFERKAREVCNDLENFTSGSPFLCMDLSTYITALLKDGFGFAD 395  
QY 420 SKVLKTRKIDNVTSMALGAI FHYIDSL 448  
DB 396 STVLQTLTKVNNIETGMALGATFHLQSL 424

## RESULT 14

US-09-370-265-5  
; Sequence 5, Application US/09370265  
; Patent No. 6447771  
; GENERAL INFORMATION:  
; APPLICANT: Ford, John  
; APPLICANT: Mulero, Jullio  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE  
; TITLE OF INVENTION: POLYPEPTIDES  
; FILE REFERENCE: 2811/35908  
; CURRENT APPLICATION NUMBER: US/09/370,265  
; EARLIER FILING DATE: 1999-08-09  
; EARLIER APPLICATION NUMBER: PCT/US99/16180  
; EARLIER FILING DATE: 1999-07-16  
; EARLIER APPLICATION NUMBER: 09/350,836  
; EARLIER FILING DATE: 1999-07-09  
; EARLIER APPLICATION NUMBER: 09/273,447  
; EARLIER FILING DATE: 1999-03-19  
; EARLIER APPLICATION NUMBER: 09/244,444  
; EARLIER FILING DATE: 1999-02-04  
; EARLIER APPLICATION NUMBER: 09/122,449  
; EARLIER FILING DATE: 1998-07-24  
; EARLIER APPLICATION NUMBER: 09/118,205  
; EARLIER FILING DATE: 1998-07-16  
; NUMBER OF SEQ ID NOS: 37  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 428  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-370-265-5

Query Match 42.3%; Score 999; DB 4; Length 428;

Best Local Similarity 52.4%; Pred. No. 2.8e-96;  
Matches 204; Conservative 56; Mismatches 123; Indels 6; Gaps 4;

QY 62 PLGTAAADGHEVYFVIMPDAGSTGTRVHVFOFT-RPPRETPLTHETFKAVKPGLSAYVD 120  
DB 40 PINVSA---STLYGIMFDPAGSTGTRIHVYTFVQKMPQGLPLBGEVFDSPKPSAFVDQ 96  
QY 121 VEKSAQGIREFLIDVAKODIPDFWKATPVLKATAGRLLPGEKAKQLLOKYKVEFKASP 180  
DB 97 PKQAEVYQGLLEVAKDISPRSHMKKTPVVKATAGRLLPENKAKALLFEVKEIFRXPSP 156  
QY 181 FLVGGDDCVSINMGDEGVSANITINFLTGSUKTGSSVGMULDGGSGSTQIAFLPRVEGT 240  
DB 157 FLVPRGVSIVMGDSDEGLAVTVNFTLGQJHGRQETVGLDGGASTQITFLPQFEKT 216  
QY 241 LQASPPGYLTALRMFNRTYKLYSYSLGLGMSARLAILGGVEGQPADGKELVSPCLSP 300  
DB 217 LEQTPRGVLTSPFEMFNSTYKLYTHSYLFGKAKARLITGLALETE-GTDGHTFRSACLPR 275  
QY 301 SFKGEWEHAETVYRVSQKAAASLHELCAARYSEVLONRVHRTVEYKVDVPAFSYYIDL 360  
DB 276 WLEAEWIFGVYKYQYGVNOGEVGEPEPCYAEVLRVVRKQLQBPBEVORGSEFYAFSYYYDR 335  
QY 361 AAGVGLIDAEKGGSLVVGDFEIAKVCRTLETQPOSSPCMDLTYVSLLOE-FGPPR 419  
DB 336 AVDTMDIDYKGGILKVEDFERKAREVCNDLENFTSGSPFLCMDLSTYITALLKDGFGFAD 395  
QY 420 SKVLKTRKIDNVTSMALGAI FHYIDSL 448  
DB 396 STVLQTLTKVNNIETGMALGATFHLQSL 424

## RESULT 15

US-09-557-800C-3  
; Sequence 3, Application US/09557800C  
; Patent No. 6476211  
; GENERAL INFORMATION:  
; APPLICANT: Ford, John

APPLICANT: Mulero, Julio  
 APPLICANT: Yeung, George  
 TITLE OF INVENTION: Methods and Materials Relating to CD39-Like  
 TITLE OF INVENTION: Polypeptides  
 FILE REFERENCE: 28110/36457  
 CURRENT APPLICATION NUMBER: US/09/557,800C  
 CURRENT FILING DATE: 2000-04-25  
 PRIOR APPLICATION NUMBER: 09/481,238  
 PRIOR FILING DATE: 2000-01-11  
 PRIOR APPLICATION NUMBER: 09/370,265  
 PRIOR FILING DATE: 1999-08-09  
 PRIOR APPLICATION NUMBER: PCT/US99/16180  
 PRIOR FILING DATE: 1999-07-16  
 PRIOR APPLICATION NUMBER: 09/350836  
 PRIOR FILING DATE: 1999-07-09  
 PRIOR APPLICATION NUMBER: 09/273447  
 PRIOR FILING DATE: 1999-03-19  
 PRIOR APPLICATION NUMBER: 09/122449  
 PRIOR FILING DATE: 1998-07-24  
 PRIOR APPLICATION NUMBER: 09/244444  
 PRIOR FILING DATE: 1999-02-04  
 PRIOR APPLICATION NUMBER: 09/118,205  
 PRIOR FILING DATE: 1998-07-16  
 NUMBER OF SEQ ID NOS: 56  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 3  
 LENGTH: 428  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-557-800C-3

Query Match 42.3%; Score 999; DB 4; Length 428;  
 Best Local Similarity 52.4%; Pred. No. 2.8e-96;  
 Matches 204; Conservative 56; Mismatches 123; Indels 5; Gaps 4;

Qy	62	PLGTADAGHEVFYGMFDAGSTCTRVHVQFT-RPPREPTTLTHETFKAVKPLGSLAYADD	120
Db	40	PINVSA---STLYGIMFDAGSTCTRIHVTFVQKMGQLFILEGEVDSVKPGLSAFVDQ	96
Qy	121	VEKSAQGIPELLDVAKODIPDFDKATPLVLKATAGLRLLPGEKACKLQKVEVFKASP	180
Db	97	PKQAEVQGLLEVAKDSIPRSHKKTPVVLKATAGLRLLPEKAKALLFEVKEIFRKSP	156
Qy	181	FLVGDDCVSIMNGTDEGVSAWITINFLTGLSKTPGSSVGMLDLGGSGTQIAFLPRVEGT	240
Db	157	FLVPKGSVIMDGSDEGILAWTVNFLTGLGHGHOETVGTLDLGGASTQITFLPQFEKT	216
Qy	241	LQASPPGYLTALSMFNETHYKLYSYSLGLGLMSARLAILGGVEGQPAKDGKELVSPCLSP	300
Db	217	LEQTPRGYLTSEFMFNSTYKLYTHSYGLGLKAARLATIGALETE-GTDGHTFRSACLPR	275
Qy	301	SFKGEWEHAEVTVRVSCQKAAASLHELCAARVSEVLQNRVHRTVEEVKHVDVFAFSYYIDL	360
Db	276	WLEAEWIFGGVKYQYGGNQGSEVGFPCYAEVLRVVRGKLHQFEEVQSGSFYAFSYYIDR	335
Qy	361	AAGVLIDAEKGSLLVVGDPFIAAKVYCRTELETQPOSSPFCMDLTYVSLILOE-FGPPR	419
Db	336	AVDTDMIDYEKGKGLKVEDFERKAREVCNLENFTSGSPFLCMDLSYITALLKDGFGFAD	395
Qy	420	SKVLKLTNRKIDNVETSWALCAIFHYIDSL	448
Db	396	STVLQITKVNNIETGWALGATFHLQSL	424

Search completed: June 8, 2004, 09:57:36  
 Job time : 23 secs

A;Accession: G72466  
A;Status: preliminary  
A;Reference number: A/2430; MOLD: 5551033; EMD: 1038258

A;Cross-references: GB:AL590842; PIDN:CAC92508.1; PID:g15981208; GSPDB:GN00175  
A;Genetics:  
A;Gene: YFO3274

Query Match 8.4%; Score 53.5; DB 2; Length 113;  
Best Local Similarity 28.9%; Pred. No. 1.1e+02;  
Matches 22; Conservative 9; Mismatches 38; Indels 7; Gaps 3;

QY 7 HGSIRVAKVAYPLGLCVGVFIYVYIKWHRATATQAFF-SITRAAPGARWGQAHSP LGT 65  
DB 4 HYIYRELLLA-PLLITSSACTHVANDSWTKDKAKHFFASAAALAAAGTAYGEHQN----- 57  
QY 66 AADGHEVFYGIIMFDAG 81  
DB 58 WSDAKSRNFGLLFSIG 73

RESULT 5  
T09627  
positive transcription activator - Klebsiella pneumoniae  
C:Species: Klebsiella pneumoniae  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
C:Accession: T09627  
F;George, A.M.; Hall, R.M.; Stokes, H.W.  
Microbiology 141, 1909-1920, 1995  
A;Title: Multidrug resistance in Klebsiella pneumoniae: a novel gene, ramA, confers a mu  
A;Reference number: Z16785; MUID:96032015; PMID:7551053  
A;Accession: T09627  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: DNA  
A;Residues: 1-113 <GEO>  
A;Cross-references: EMBL:U19581; NID:9885876; PIDN:AAA85697.1; PID:9885878  
A;Experimental source: strain ECL8Mdr  
C;Genetics:  
A;Gene: ramA  
C;Keywords: DNA binding; transcription regulation

Query Match 8.3%; Score 53; DB 2; Length 113;  
Best Local Similarity 26.0%; Pred. No. 1.3e+02;  
Matches 25; Conservative 9; Mismatches 42; Indels 20; Gaps 4;

QY 7 HGSIRVAKVAYPLGLCVGVFIYVYIKWHRATATQAFFSITRAAPGARWGQAHSP LGT- 65  
DB 20 HQPIRIDDIARHAG-----YKWH---LQELFLQYKGESLGRVIRERKLLLAAR 65  
QY 66 -AADGHEVFYGIIM- ---FDAGSTGTRHVVFQFTRPP 96  
DB 66 DLRDTDQRVYDICKYGFDSQQQFTRVFRTRFNQPP 101

RESULT 6  
A87029  
hypothetical protein [imported] - Mycobacterium leprae  
C:Species: Mycobacterium leprae  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C;Accession: A87029  
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,  
eam, M.A.; Rutherford, K.M.  
Nature 409, 1007-1011, 2001  
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq  
A;Title: Massive gene decay in the leprosy bacillus.  
A;Reference number: A86909; MUID:21128732; PMID:11234002  
A;Accession: A87029  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-117 <SPTO>  
A;Cross-references: GB:AL450380; NID:g13093009; PIDN:CAC31340.1; GSPDB:GN00147  
C;Genetics:  
A;Gene: ML0959

Query Match 8.2%; Score 52.5; DB 2; Length 117;  
Best Local Similarity 25.0%; Pred. No. 1.5e+02;

Matches	19, Conservative	10, Mismatches	26, Indels	21, Gaps	4,
QY	3	KINHGSLRVAKVAPPLGLCVGFIVYAYIKMHRBATQAFPTTPAAGARWGQAKHP	62		
Db	47	RVDGHSRVLVTMKSL-----IRVAV-WRRPAL-----AVPLPQRESRRDP	89		
QY	63	L-----GTRADGHEVFY	74		
Db	90	AMTMDETVASHGQTFW	105		

RESULT 7  
 G95324  
 Probable Trad conjugal transfer protein [imported] - *Sinorhizobium meliloti* (strain 1021)  
 C:Species: *Sinorhizobium meliloti*  
 C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
 C:Accession: G95324  
 R:Barrett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bower  
 proc. Natl. Acad. Sci. U.S.A. 98, 9683-9688, 2001  
 A:Title: Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti*  
 A:Reference number: A95262; MUID:21396509; PMID:11481432  
 A:Accession: G95324  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-79 <KB>  
 A:Cross-references: GS:AB006469; PIDN:PAK651c1.1; PID:G14523604; GSPDB:GN00165  
 A:Experimental source: Strain 1021, megaplasmid pSymA  
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
 pela, D.; Chait, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
 L.; Hyman, R.W.; Jones, T.  
 Science 293, 668-672, 2001  
 A:Authors: Kahn, D.; Kahn, M.L.; Kallman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
 hebaull, P.; Vandenbol, M.; Vortroler, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yen, K.  
 A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.  
 A:Reference number: A96039; MUID:21368234; PMID:11474104  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: trad  
 A:Genome: plasmid

Query Match	8.0%	Score 51.5	DB 2	Length 79
Best Local Similarity	27.4%	Pred. No. 1.3e+02		
Matches 20	Conservative 6	Mismatches 38	Indels 9	Gaps 3

QY	13	A K A V A Y P E L G L C V G F I Y A A Y I K M R A T A T Q A F S I T R A A P A R W G C O A H S P L G T A A A G H E V	72
	:   :   :   :   :   :		
Db	9	A R E K F R L G --- A I V V R A G L T K D R A F L L G S F E L A V T P S --- A H R R L R N D I -- G E L A	59
QY	73	P Y G I M P D A G S T G T	85
Db	60	F K A P A L D G S S P G T	72

RESULT 8  
S58803  
ubiquinol-cytochrome-c reductase (EC 1.10.2.2) cytochrome b - *Doronomyxex kutleri* mtoc  
C/Species: mitochondrion *Doronomyxex kutleri*  
C/Date: 19-Mar-1997 #sequence\_revision 24-Oct-1997 #text\_change 03-Jun-2002  
C/Accession: S58803  
R/Baur, A.; Chalwatzis, N.; Buechinger, A.; Zimmermann, F. K.  
Curr. Genet. 28, 242-247, 1995  
A/Title: Mitochondrial DNA sequences reveal close relationships between social parasitoid  
A/Reference number: S58798; MUID:96071193; PMID:8528270  
A/Accession: S58803  
A/Molecule type: DNA  
A/Residues: 1-110 <BAU>  
C/Genetics:  
A/Genome: mitochondrion  
C/Superfamily: cytochrome b; cytochrome b homology; cytochrome b6 homology; plastocyanin  
C/Keywords: chromoprotein; electron transfer; heme; iron; mitochondrion; oxidative phosph  
E/1-110/Domain: cytochrome b homology (fragment) <CB6>  
/1-110/Domain: cytochrome b homology (fragment) <CB6>

F:9-25/Domain: transmembrane #status predicted <TM1>  
F:54-72/Domain: transmembrane #status predicted <TM2>  
F:90-106/Domain: transmembrane #status predicted <TM3>

Query Match 5.0%; Score 51.5; DB 2; Length 110;  
Best Local Similarity 29.3%; Pred. No. 1 8e+02;  
Matches 24; Conservative 10; Mismatches 35; Indels 13; Gaps 4.

QY 5 SNHSLRVAKAYAPPLCVGVFIYVA-YIKMRATATQ-AFFESTRLAAGARWGOAASP 62

Db 4 SNYGL-----LGLFLAIQIFSGFPLSIHYCPNTQLAFFSIHIMGNISNGMMHN- 54

```
QY      63 LGTADGHEVFYGYIMFDAGSTG 84
      : | : | : |
Db      55 --IHNGASFFFCMYIHMSRG 74
```

## RESULT 5

T-cell receptor beta chain precursor V region (Mlp2) - mouse (fragment)  
 C:Species: Mus musculus (house mouse)  
 C:Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 23-Jul-1999  
 C:Accession: A25903  
 R:Singer, P.A.; McGivilly, R.J.; Noonan, D.J.; Dixon, F.J.; Theofilopoulos, A.N.  
 Proc. Natl. Acad. Sci. U.S.A. 83, 7018-7022, 1986  
 A:Title: Clonal diversity and T-cell receptor beta-chain variable gene expression in en  
 A:Reference numbers: A25903; MUID:66313702; PMID:3462739  
 A:Accession: A25903  
 A:Molecule type: mRNA  
 A:Residues: 1-81 <SI>  
 A:Cross-references: GB:M14294; NID:g201349; PIDN:AAA0241.1; PID:g201350  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: T-cell receptor  
 F:1-62/Domain: V segment (fragment) <VAR>  
 F:63-64/Domain: D segment <DIV>  
 F:65-81/Domain: J segment <JOI>

Query Match	8.0%;	Score 51;	DB 2;	Length 81;
Best Local Similarity	29.2%;	Pred. NC. 1.5e+02;		
Matches 19;	Conservative 12;	Mismatches 24;	Indels 10;	Gaps 3

QY 52 GAATGAGCOOAAHPICLTGAADGHEVAFVGIMFGAGSTGRVHVFQFTPRPREFTLTHETPKAVK 111  
 : : : : : : : : : : : : : : : :  
 Db 2 GAATFOEGSLMTAATANEGSEATY---ESGFTKD---PPISRNLTBSTL---TVNNAR 51

Qy	112	PGLSA	116
			:
Db	52	PGDSS	56

RESULT 10  
PQ0115  
hypothetical protein (azu region) - *Pseudomonas aeruginosa* (fragment)

C|Date: 31-Mar-1992 #sequence\_revision 31-Dec-1993 #text\_change 08-Oct-1999  
C|Accession: RQ0115  
R|Hotlink, C.W.G.; Wouda, L.P.; Threnhout, J.C.M.; van de Kamp, M.; Canters, G.W.  
Gene 90, 15-20, 1990  
A|Title: Isolation and sequencing of the Alcaligenes denitrificans azurin-encoding gene.  
A|Reference number: JQ0643; MUID:90337337; PMID:2116366  
A|Accession: RQ0115  
A|Molecule type: DNA  
A|Residues: 1-100 <HOT>  
A|Cross-References: GB:J030389, NID:g151060, PIDN:AAA2731.1, PID:g151063

Query March	7.9%	Score	50.5;	DB	2;	Length	100;
Best Local Similarity		Pred. No.	2.1e+02;				
Matches	13;	Conservative	4;	Mismatches	16;	Indels	5;
						Gaps	1;

QY 22 CVCVFIVAYIKWH-----PATATQAFFSITRLAAGAR 54  
| : : : : :  
Db 62 CANPDQYFWDEWHPTRRVHQLAGEMAAPTRAELGAR 99



```
RESULT 11
F70909
hypothetical protein Rv0603 - Mycobacterium tuberculosis (strain H37RV)
C/Species: Mycobacterium tuberculosis
C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 28-Jul-2000
C/Accession: F70909
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellon, S.; Squares, S.
Nature 393, 537-544, 1998
A/Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A/Reference number: A70500; MUID:98295987; PMID:9634230
A/Accession: F70909
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-103 <COL>
A/Cross-references: GB:297182; GB:AL123456; NID:G3250720; PIDN:CA809954.1; PID:e3241175;
A/Experimental source: strain H37RV
C/Genetics:
A/Gene: Rv0603
C/Superfamily: Mycobacterium tuberculosis hypothetical protein Rv0603
Query Match 7.9%; Score 50.5; DB 2; Length 103;
Best Local Similarity 29.0%; Pred. No. 2.2e+02;
Matches 27; Conservative 12; Mismatches 41; Indels 13; Gaps 5;
QY 1 MRKISNHSRLRVAKVAPLGLCVGVFIYVAYIKWHEAT---ATQAFFSITRAAPGARWGO 57
DB 1 MNRIVQFVSNAVAAA--TGINGSGIAAFAFGEDVETGPDRAARAQAQVFGGTAGE 58
QY 58 -QAHSPGLTAADGHEVYFIMFDAGSTGRVHV 89
DB 59 VETETGEGAAA-----YGLVLR-TRPDGTRREV 84
RESULT 12
T13211
minor capsid protein R117b - Lactobacillus phage phi-gle
C/Species: Lactobacillus phage phi-gle
C/Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 21-Jul-2000
C/Accession: T13211
R/Kodaira, K.I.; Oki, M.; Kakikawa, M.; Watanabe, N.; Hirakawa, M.; Yamada, K.; Taketo,
Gene 187, 45-53, 1997
A/Title: Genome structure of the Lactobacillus temperate phage phi gle: the whole genome
A/Reference number: Z17631; MUID:97225795; PMID:9073065
A/Accession: T13211
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-116 <KOD>
A/Cross-references: EMBL:X98106; NID:G1926320; PIDN:CAA66740.1; PID:G1926355
C/Genetics:
A/Note: Ror117b
Query Match 7.8%; Score 50; DB 2; Length 116;
Best Local Similarity 22.6%; Pred. No. 2.8e+02;
Matches 21; Conservative 13; Mismatches 39; Indels 20; Gaps 3;
QY 37 ATATQAFFSITRAAPGARWGOQAHSPGLTAADGHEV-----FYGIMFDAGSTGTR 86
DB 27 ALVNQAMYNDEQFVFKDRPEELRQSVHATSDGSEITYSTPYAKAQFYGIIND-----K 80
QY 87 VHVFOFTRPPRPTLTTHETFKAVKGLSAYAD 119
DB 81 YPVHNYTTP----GTTKRWDLKAKSMFMSWID 109
RESULT 13
S38768
rho protein GDP dissociation inhibitor - guinea pig (fragments)
C/Species: Cavia porcellus (guinea pig)
C/Date: 09-Jun-1994 #sequence_revision 06-Sep-1996 #text_change 18-Aug-2000
```

```
C/Accession: S38768
R/Pick, E.; Gorzalczany, Y.; Engel, S.
Eur. J. Biochem. 217, 441-455, 1993
A/Title: Role of the rac1 p21-GDP-dissociation inhibitor for rho heterodimer in the acti
A/Reference number: S38767; MUID:94039069; PMID:8223583
A/Accession: S38768
A/Molecule type: protein
A/Residues: 1-40; 41-46; 47-86; 87-106 <PIC>
C/Superfamily: human rho GDP dissociation inhibitor
Query Match 7.7%; Score 49.5; DB 2; Length 106;
Best Local Similarity 29.3%; Pred. No. 2.9e+02;
Matches 12; Conservative 4; Mismatches 20; Indels 5; Gaps 1;
QY 70 HEVFGYGMFD-----AGSTGTRVHVFOFTRPPRPTLTTH 105
DB 50 HTRKGVKIDTYMVGSGFRAEEYEFLLTPMEAPKPTDD 90
RESULT 14
T06971
hypothetical protein 32 - Cyanophora paradoxa cyanelle
C/Species: cyanelle Cyanophora paradoxa
C/Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 12-Jun-2003
C/Accession: T06971
R/Stirewalt, V.L.; Michalowski, C.B.; Luffelhardt, W.; Bohnert, H.J.; Bryant, D.A.
submitted to the EMBL Data Library, July 1995
A/Description: Nucleotide sequence of the cyanelle genome from Cyanophora paradoxa.
A/Reference number: Z15840
A/Accession: T06971
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-38 <STI>
A/Cross-references: EMBL:U30821; NID:G1016083; PIDN:AAA81314.1; PID:G1016227
A/Experimental source: strain Pringsheim LB555
C/Genetics:
A/Gene: ycf32
A/Genome: cyanelle
C/Superfamily: uncharacterized conserved protein ycf32
C/Keywords: cyanelle
Query Match 7.7%; Score 49; DB 2; Length 38;
Best Local Similarity 34.4%; Pred. No. 1.1e+02;
Matches 11; Conservative 7; Mismatches 10; Indels 4; Gaps 1;
QY 9 SLRVAKVAYPLGLCVGVFIY----VAYIKWHR 36
DB 2 SMRLVVVLLPLGLGALGWAVYNGKLAIEQWRR 33
RESULT 15
GMHUB
gastrin precursor [validated] - human
N/Contains: big gastrin; cryptogastrin; gastrin; gastrin-17
C/Species: Homo sapiens (man)
C/Date: 24-Apr-1984 #sequence_revision 30-Jun-1987 #text_change 19-Jan-2001
C/Accession: A93497; A93497; A94473; A93152; A91628; A18854; A40869; A32487; C32
R/Ito, R.; Sato, K.; Helmer, T.; Jay, G.; Agarwal, K.
Proc. Natl. Acad. Sci. U.S.A. 81, 4662-4666, 1984
A/Title: Structural analysis of the gene encoding human gastrin: the large intron contai
A/Reference number: A93997; MUID:84272693; PMID:6087340
A/Accession: A93997
A/Molecule type: DNA
A/Residues: 1-101 <ITO>
A/Cross-references: GB:K01254; GB:J00147; NID:G182987; PIDN:AAB59533.1; PID:G182989
R/Kato, K.; Hayashizaki, Y.; Takahashi, Y.; Himeno, S.; Matsuura, K.
Nucleic Acids Res. 11, 8197-8203, 1983
A/Title: Molecular cloning of the human gastrin gene.
A/Reference number: A93497; MUID:84169471; PMID:6324077
A/Accession: A93497
A/Molecule type: DNA
A/Residues: 1-101 <KAT>
A/Cross-references: GB:X00183; NID:G31648; PIDN:CAA25005.1; PID:G31649
```

R.Harris, J.I.; Kenner, E.W.  
 unpublished results, cited by Gregory, R.A., and Tracy, H.J., in Gastrointestinal Hormon  
 A:Reference number: A94473  
 A:Accession: A94473  
 A:Molecule type: protein  
 A:Residues: 59-64, 'H', '66', 'S', '68-92' <HAR>  
 A:Experimental source: Gastrinoma tissue  
 R.Bentley, P.H.; Kenner, G.W.; Sheppard, R.C.  
 Nature 209, 583-585, 1966  
 A:Title: Human gastrin isolation, structure and synthesis.  
 A:Reference number: A93152; PMID:67021327; PMID:5921183  
 A:Accession: A93152  
 A:Molecule type: protein  
 A:Residues: 76-92 <BEN>  
 A:Note: Gastrin comprises the carboxyl-terminal 17 residues of big gastrin  
 R.Gregory, R.A.; Tracy, H.J.; Agarwal, K.L.; Grossman, M.I.  
 Gut 10, 603-608, 1969  
 A:Title: Amino acid constitution of two gastrins isolated from Zollinger-Ellison tumour  
 A:Reference number: A91628; PMID:69298172; PMID:5822140  
 A:Accession: A91628  
 A:Molecule type: protein  
 A:Residues: 76-92 <GRE>  
 A:Note: the gastrin was isolated from human Zollinger-Ellison pancreatic tumor  
 R.Wiborg, O.; Berglund, L.; Boel, E.; Norris, P.; Rehfeld, J.F.; Marcker, K.  
 Proc. Natl. Acad. Sci. U.S.A. 81, 1067-1069, 1984  
 A:Title: Structure of a human gastrin gene.  
 A:Reference number: A18854; PMID:84144842; PMID:6322186  
 A:Accession: A18854  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-101 <MIB>  
 A:Cross-references: GB:K01254; GB:J00147; NID:G182967; PIND:AA59533.1; PTD:G182989  
 R.Hubner, V.D.; Jiang, R.; Lee, T.D.; Legesse, K.; Walsh, J.H.; Shively, J.E.; Chew, P.  
 J. Biol. Chem. 266, 12223-12227, 1991  
 A:Title: Purification and structural characterization of progastrin-derived peptides frc  
 A:Reference number: A40869; PMID:91286536; PMID:2061307  
 A:Accession: A40869  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 22-51 <HUB>  
 R.Higashimoto, Y.; Himeno, S.; Shinomura, Y.; Nagao, K.; Tamura, T.; Tani, S.  
 Biochem. Biophys. Res. Commun. 160, 1364-1370, 1989  
 A:Title: Purification and structural determination of urinary NH-2-terminal big gastrin  
 A:Reference number: A32487; PMID:89273602; PMID:2730647  
 A:Accession: A32487  
 A:Molecule type: protein  
 A:Residues: 59-67 <H12>  
 A:Note: this urinary fragment of big gastrin was designated peak Ia  
 A:Experimental source: urine  
 A:Accession: C32487  
 A:Molecule type: protein  
 A:Residues: 59-66 <H13>  
 A:Note: this urinary fragment of big gastrin was designated peak Ib  
 A:Experimental source: urine  
 A:Accession: C32487  
 A:Molecule type: protein  
 A:Residues: 59-68 <H14>  
 A:Note: this urinary fragment of big gastrin was designated peak II  
 R.Higashimoto, Y.; Shinomura, Y.; Nagao, K.; Yasunaga, Y.; Tamura, T.; Tani, S.  
 Biochem. Biophys. Res. Commun. 172, 1392-1399, 1990  
 A:Title: Purification of N-terminal hexapeptide of big gastrin from human urine.  
 A:Reference number: A36249; PMID:91058586; PMID:2244919  
 A:Accession: A36249  
 A:Molecule type: protein  
 A:Residues: 59-64 <H15>  
 A:Note: this urinary fragment of big gastrin was designated peak III  
 R.Boel, E.; Vunst, J.; Norris, P.; Rehfeld, J.F.; Marcker, K.A.  
 Proc. Natl. Acad. Sci. U.S.A. 80, 2866-2869, 1983  
 A:Title: Molecular cloning of human gastrin cDNA: evidence for evolution of gastrin by g  
 A:Reference number: 137408; PMID:83221503; PMID:6574456  
 A:Accession: 137408  
 A:Molecule type: mRNA

A:Residues: 1-101 <RES>  
 A:Cross-references: EMBL:V00511; NID:G31654; PIND:CAA23769.1; PID:G31655  
 R.Rehfeld, J.F.; Johnsen, A.H.  
 Eur. J. Biochem. 223, 765-773, 1994  
 A:Title: Identification of gastrin component I as gastrin-71. The largest possible bloac  
 A:Reference number: S48183; PMID:94533379; PMID:8055952  
 A:Accession: S48183  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 22-40 <REH>  
 R.Kariya, Y.; Kato, K.; Hayashizaki, Y.; Himeno, S.; Tani, S.; Matsubara, K.  
 Gene 50, 345-352, 1986  
 A:Title: Expression of human gastrin gene in normal and gastrinoma tissues.  
 A:Reference number: 154006; PMID:87219893; PMID:3034736  
 A:Accession: 154006  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-101 <KAR>  
 A:Cross-references: GB:M15958; NID:G182990; PIND:AAA52520.1; PTD:G182991  
 R.Rehfeld, J.F.; Hansen, C.P.; Johnsen, A.H.  
 EMBO J. 14, 389-396, 1995  
 A:Title: Post-poly(Glu) cleavage and degradation modified by O-sulfated tyrosine: a nove  
 A:Reference number: S54350; PMID:95137019; PMID:7530658  
 A:Accession: S54350  
 A:Molecule type: protein  
 A:Residues: 76-92 <KEM>  
 A:Note: gastrin-6 isolated from gastric mucosa and blood is fully active; sulfation of  
 C:Genetics:  
 A:Gene: GDB:GAS  
 A:Cross-references: GDB:119261; OMIM:137250  
 A:Map position: 17q-17q  
 A:Insertions: 71/1  
 C:Superfamily: Gastrin  
 C:Keyword: amidated carboxyl end; phosphoprotein; pyroglutamic acid; sulfoxprotein  
 F:1/21/Domain: signal sequence #status predicted <SIG>  
 F:2/25/Product: cryptogastrin (amino-terminal propeptide) #status experimental <PRO>  
 F:59-92/Product: big gastrin #status experimental <BGN>  
 F:76-92/Product: gastrin #status experimental <SGN>  
 F:87-92/Product: gastrin-6 #status experimental <GN6>  
 F:55/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #link BGN #status  
 F:76/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #link SGN #status  
 F:87/Binding site: sulfate (Tyr) (covalent) (partial) #status experimental  
 F:92/Modified site: amidated carboxyl end (Phe) (amide in mature form from following Gly  
 Query Match 7.7%; Score 49; DB 1; Length 101;  
 Best Local Similarity 35.3%; Pred. No. 3, 1e+02;  
 Matches 18; Conservative 3; Mismatches 12; Indels 18; Gaps 3;  
 QY 21 LCYGVFIYVAYIKMHRATATCAFFSITRAAPARM---GQAHSPILGTAAD 68  
 Db 4 LCYVVLIF-----ALAAAFSE-----ASMKPRSQQDADPLGTGAN 39

Search completed: June 8, 2004, 10:02:12  
 Job time : 22 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 09:57:10 ; Search time 17 Seconds  
(without alignments)  
367.554 Million cell updates/sec

Title: US-09-905-589A-2\_COPY\_1\_120

Perfect score: 640

Sequence: 1 MRKSNHGSRLVAKVAPLG.....TLHETPKAVKPGLSAYADD 120

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 20892

Minimum DB seq length: 0

Maximum DB seq length: 120

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	8.3	113	1	RAMA_KLEPN
2	52	8.1	113	1	RAMA_KLEPN
3	51	8.0	101	1	PER2_MOUSE
4	49.5	7.7	117	1	PAND_BRAVA
5	49	7.7	38	1	PSBY_CYAPA
6	49	7.7	101	1	GAST_HUMAN
7	48.5	7.6	107	1	VEA_BPV4
8	48.5	7.6	111	1	GDIF_CAVPO
9	48	7.5	83	1	PSBE_CHLVU
10	47.5	7.4	83	1	VFP_BPRD
11	47.5	7.4	115	1	VPM_BPR2
12	47.5	7.4	115	1	YC9C_SCHPO
13	47	7.3	106	1	YAHM_ECOLI
14	47	7.3	116	1	NUJM_CYPCA
15	46.5	7.3	97	1	Y109_EMENT
16	46.5	7.3	106	1	YNE3_YEAST
17	46.5	7.3	118	1	RL7_MICLU
18	46	7.2	76	1	NIKY_BOVIN
19	46	7.2	95	1	Y736_ARCFU
20	45.5	7.1	63	1	YPH2_SYNP2
21	45.5	7.1	102	1	TRAC_RHISN
22	45.5	7.1	106	1	ONCB_TAETA
23	45	7.0	83	1	PSBE_NEPOL
24	45	7.0	101	1	PER2_HUMAN
25	45	7.0	109	1	YVDJ_BACUS
26	45	7.0	116	1	BRJ1_BRARE
27	44.5	7.0	85	1	CRH_BACSV
28	44.5	7.0	103	1	FER_SULAC
29	44.5	7.0	113	1	Y95J_CAMJE
30	44	6.9	67	1	CERC_CERCA
31	44	6.9	82	1	PSBE_CHIRE
32	44	6.9	94	1	Y51B_METUA
33	44	6.9	111	1	LV2H_HUMAN

#### ALIGNMENTS

34	44	6.9	111	1	YHB2_STRCO	P23159 streptomyc
35	44	6.9	116	1	PO41_CHICK	P55968 gallus gall
36	44	6.9	118	1	YVBI_HUMAN	Q13541 homo sapien
37	43.5	6.8	67	1	YVBI_VACCC	P20549 vaccinia vi
38	43.5	6.8	80	1	R318_STRMU	Q84tn5 streptococc
39	43.5	6.8	82	1	P232_MERIN	Q99pe3 meriones un
40	43.5	6.8	98	1	EXOX_RHIME	Q02730 rhizobium m
41	43.5	6.8	109	1	PT16_STYPL	P28208 styela pluc
42	43.5	6.8	111	1	Y5ND_BACSV	P94563 bacillus su
43	43.5	6.8	117	1	DHA_CRAPL	P22238 craterostig
44	43.5	6.8	117	1	HV10_MOUSE	P01754 mus musculu
45	43.5	6.8	117	1	RBPA_LACPL	Q88vk8 lactobacill

#### RESULT 1

ID RAMA\_KLEPN STANDARD; PRT; 113 AA.

AC Q48413;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DE Transcriptional activator ramA.  
GN RAMA.  
OS Klebsiella pneumoniae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Klebsiella.  
OX NCBI\_TaxID=573;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=6032015; PubMed=7551053;  
RX STRAIN=EC18;  
RA George A.M., Hall R.M., Stokes H.W.;  
RT "Multidrug resistance in Klebsiella pneumoniae: a novel gene, ramA,  
confers a multidrug resistance phenotype in Escherichia coli.";  
RL Microbiology 141:1909-1920(1995).  
CC -!- FUNCTION: PROBABLE TRANSCRIPTIONAL ACTIVATOR. CONFERS A MULTIDRUG  
RESISTANCE PHENOTYPE IN ESCHERICHIA COLI.  
CC -!- SIMILARITY: BELONGS TO THE ARAC/XYLIS FAMILY OF TRANSCRIPTIONAL  
REGULATORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U19581; AAA85697.1; -.  
DR PIR; T03627; T09627.  
DR HSSP; P27246; 1BL0.  
DR InterPro; IPR000005; HTHARAC.  
DR Pfam; PF00165; HTH\_ARAC; 2.  
DR PRINTS; PR00032; HTHARAC.  
DR SMART; SM00342; HTH\_ARAC; 1.  
DR PROSITE; PS00041; HTH\_ARAC\_FAMILY\_1; 1.  
DR PROSITE; PS01124; HTH\_ARAC\_FAMILY\_2; 1.  
KW Transcription regulation; DNA-binding; Activator.  
FT DNA BIND 25 44 H-T-H motif (POTENTIAL).  
SQ SEQUENCE 113 AA; 13495 MW; 46E0D415C066C9F CRC64;

Query Match 8.3%; Score 53; DB 1; Length 113;

Best Local Similarity 26.0%; Pred. No. 68;

Matches 25; Conservative 9; Mismatches 42; Indels 20; Gaps 4;

QY 7 HGSRLVAKVAPLGCVGIVIVATIKYKRAATQAFPTMAAGARWGQQAHSPLG- 65  
DB 20 HGPLRIDDIARHAG-----YSKW---LQRLFYQYGSISLGRYIRERKILLAR 65  
QY 66 -AADGHEVEFYGIM---FDAGSTGTRVHVFQETRP 96

Db 66 DLRTDQVYDCLKGFDSQSTFTVFTFNQPP 101  
RESULT 2  
RAMA\_ENTCL  
ID RAMA\_ENTCL STANDARD; - PRT; 113 AA.  
AC P5922;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DE 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Transcriptional activator ramA.  
GN RAMA.  
OS Enterobacter cloacae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Enterobacter.  
OX NCBI\_TaxID=550;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90299838; PubMed=2193928;  
RA Komatsu T., Ohta M., Kido N., Azakawa Y., Ito H., Mizuno T., Kato N.;  
RT "Molecular characterization of an Enterobacter cloacae gene (romA)  
RT which pleiotropically inhibits the expression of Escherichia coli  
RT outer membrane proteins.";  
RL J. Bacteriol. 172:4082-4089(1990).  
RN [2]  
RP IDENTIFICATION.  
RX MEDLINE=96032015; PubMed=7551053;  
RA George A.M., Hall R.M., Stokes H.W.;  
RT "Multidrug resistance in Klebsiella pneumoniae: a novel gene, ramA,  
RT confers a multidrug resistance phenotype in Escherichia coli.";  
RL Microbiology 141:1909-1920(1995).  
CC -!- FUNCTION: Probable transcriptional activator.  
CC -!- SIMILARITY: BELONGS TO THE ARAC/XYL5 FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.  
DR HSP; 272246; 1BL0.  
DR InterPro: IPR000005; HTHArac.  
DR Pfam: PF00165; HTH\_Arac; 2.  
DR PRINTS: PR00032; HTHARAC.  
DR SMART: SM00342; HTH\_ARAC; 1.  
DR PROSITE; PS00041; HTH\_ARAC\_FAMILY\_1; FALSE NEG.  
DR PROSITE; PS01124; HTH\_ARAC\_FAMILY\_2; 1.  
KW Transcription regulation; DNA-binding; Activator.  
FT DNA BIND 25 44 H-T-H MOTIF (POTENTIAL).  
SQ SEQUENCE 113 AA; 13633 MW; D9D57E77D64B9557 CRC64;  
Query Match 8.1%; Score 52; DB 1; Length 113;  
Best Local Similarity 24.1%; Pred. No. 88;  
Matches 26; Conservative 13; Mismatches 45; Indels 24; Gaps 5;  
Qy 7 HGSRLVAKVPLGLGVFIYVAYIKWHRATATQAFSITRAAPGARWGOQAHSLPGT- 65  
Db 20 HQPLRIEDIAHAG-----YSKWH-----LQRLFLQYKGSIGRYIREKLLLAAR 65  
Qy 66 -AAGCHEVFYGM----FDAGSTGTRVHVVFQTRPP-----RETFTLTH 104  
Db 66 DLRSDDRVVEICLRYGFSQQTFTFTFTFQPPGAYKVENHSRTH 113  
RESULT 3  
PEN2\_MOUSE  
ID PEN2\_MOUSE STANDARD; PRT; 101 AA.  
AC Q9C0R7;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DE 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Gamma-secretase subunit PEN-2 (Presenilin enhancer protein 2).  
GN PEN2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Kidney, Pancreas, and Testis;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schiraldi L.M., Stauber P., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Kono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kottukuri S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Colon;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strauberg R.L., Feingold S.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Dapchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Prange C.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Schaefer C.F.,  
RA Raha S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Woxley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Farley J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bonfield G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -!- FUNCTION: Essential subunit of the gamma-secretase complex. an  
CC endoprotease complex that catalyzes the intramembrane cleavage of  
CC integral membrane proteins such as Notch receptors and APP (beta-  
CC amyloid precursor protein). Probably represents the last step of  
CC maturation of gamma-secretase, facilitating endoproteolysis of  
CC presenilin and conferring gamma-secretase activity (By similarity).  
CC -!- SUBUNIT: Component of the gamma-secretase complex, a complex  
CC composed of a presenilin homodimer (PSEN1 or PSEN2), nicastrin  
CC (NCSTN), APH1 (APH1A or APH1B) and PEN2. Such minimal complex is  
CC sufficient for secretase activity, although other components may  
CC exist (By similarity).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC Predominantly located in the endoplasmic reticulum and in the cis-  
CC Golgi (By similarity).  
CC -!- SIMILARITY: Belongs to the PEN-2 family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AK002440; BAB22102.1; -.  
CC EMBL; AK006284; BAB24503.1; -.

```

DR EMBL AK007621: BAB25141.1: -.
DR EMBL BC024347: AAH24347.1: -.
DR MGI:1913590: 1700023M9Rik.
KM Transmembrane; Golgi stack; Endoplasmic reticulum.
FT DOMAIN 1 17 LUMENAL (POTENTIAL).
FT TRANSMEM 18 38 POTENTIAL.
FT DOMAIN 39 60 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 61 81 POTENTIAL.
FT DOMAIN 82 101 LUMENAL (POTENTIAL).
SQ SEQUENCE 101 AA; 11999 MW; 6596CA2984FBCC59 CRC64;

Query Match 8.0%; Score 51; DB 1; Length 101;
Best Local Similarity 23.2%; Pred. No. 1e+02;
Matches 22; Conservative 15; Mismatches 44; Indels 14; Gaps 3;

QY 1 MKKSNHSLRYAKVAPLGLGVGVIVVATIKMHRATATQAFSITRAAGAGAGGQAH 60
DB 3 LERVNNEKLNCRKRYIGGFAPLPLVNLVFW-----FTRAFIAPAYTEGQIK 54
QY 61 SPLGTADGHEVYGYIMFDAGSTGTIRVHVFQTRP 95
DB 55 GYVMSAVGF-LFWIIL-----ATWITIFQIRP 83

RESULT 4
PAND_BRAJA STANDARD; PRT; 117 AA.
ID PAND_BRAJA
AC Q9AMS6;
DT 16-OCT-2001 (Rel. 40; Created)
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 10-OCT-2003 (Rel. 42; Last annotation update)
DE Aspartate 1-decarboxylase precursor (EC 4.1.1.11) (Aspartate alpha-decarboxylase).
OS PAND OR ID899 OR BLR2096.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxId=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110spc4;
RX MEDLINE=21101824; PubMed=1157954;
RA Goettfert M., Roethlisberger S., Kuendig C., Beck C., Marty R., Henecke H.;
RT "Potential symbiosis-specific genes uncovered by sequencing a 410-kb DNA region of the Bradyrhizobium japonicum chromosome.";
RL J. Bacteriol. 183:1405-1412 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=2248498; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiyumi T., Saasomoto S., Ideasa K., Iritani M., Kawashima K., Kohara M., Matsunoto M., Shimpō S., Tsurutoka H., Wada T., Yamada M., Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197 (2002).
RN [3]
RP CARBOLYTIC ACTIVITY: L-aspartate = beta-alanine + CO(2).
CC -1- COFACTOR: Pyruvoyl group (by similarity).
CC -1- PATHWAY: Pantothenate biosynthesis; second branch.
CC -1- SIMILARITY: Belongs to the pand family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.ebi-sib.ch/announce/ or send an email to license@sib-sib.ch).
DR EMBL AF323013: AAG61071.1: -.
DR EMBL AP005942: BAC47361.1: -.

```

```

DR HSPB P31664: IAMB.
DR HAMAP: MF 00446: -1.
DR InterPro: IPR009010; Asp_decarb_fold.
DR InterPro: IPR003150; Asp_decarb.
DR Pfam: PF02261; Asp_decarb.1.
DR ProDom: PD009294; Asp_decarb.1.
DR TIGRFAMs: TIGR00223; pand; 1.
KM Pantothenate biosynthesis; Lysase; Decarboxylase; Pyruvate; Zymogen; Complete proteome.
FT CHAIN 1 24 ASPARTATE 1-DECARBOXYLASE BETA CHAIN (BY SIMILARITY).
FT CHAIN 25 117 ASPARTATE 1-DECARBOXYLASE ALPHA CHAIN (BY SIMILARITY).
FT MOD_RES 25 25 CONVERTED TO A PYRUVYL GROUP (BY SIMILARITY).
SQ SEQUENCE 117 AA; 12895 MW; B468238A20B50RCE CRC64;

Query Match 7.7%; Score 49.5; DB 1; Length 117;
Best Local Similarity 26.8%; Pred. No. 1.7e+02;
Matches 26; Conservative 8; Mismatches 38; Indels 25; Gaps 5;

QY 33 KMRATATQA-----FSITRAAPGARWQQAHSPLGTADGHEVYGYIMFDAGSTGR 86
DB 9 KHRASVTADLHYBSISIDRTLEA-----AGVIVRERIVNV-----ETGR 54
QY 87 VHVGFTRPPRE-TPTLHETFRVAVPG---LSAVA 118
DB 55 FATVYIEAPPMSGTMSLNGAARLVWPGDITIVAVA 91

RESULT 5
PSBY_CYAPA STANDARD; PRT; 38 AA.
ID PSBY_CYAPA
AC P48272;
DT 01-FEB-1996 (Rel. 33; Created)
DT 01-FEB-1996 (Rel. 33; Last sequence update)
DT 10-OCT-2003 (Rel. 42; Last annotation update)
DE Photosystem II protein Y.
OS PSBY OR YCF32.
OS Cyanophora paradoxa.
OC Cyanella.
OC Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
OX NCBI_TaxId=2762;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RA Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J., Bryant D.A.;
RT "Nucleotide sequence of the cyanella DNA from Cyanophora paradoxa";
RL Plant Mol. Biol. Rep. 13:327-332 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M., Farley J.Y., Schluchter W.M., Chung S., Neumann-Spallart C., Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;
RT "The complete sequence of the cyanella genome of Cyanophora paradoxa: the genetic complexity of a primitive plastid.";
RL (in) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E., Schwemmler W. (eds.);
RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg (1997).
RN [3]
RP FUNCTION: Manganese-binding polypeptide with L-arginine metabolizing enzyme activity. Component of the core of photosystem II (by similarity).
CC -1- SUBCELLULAR LOCATION: Cyanella thylakoid membrane.
CC -1- SIMILARITY: Belongs to the psby family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U30821; AAA1314.1; -;  
 DR PIR; T06971; T06971.1;  
 DR HANAP; MF 00717; -; 1.  
 KW Photosynthesis; Photosystem II; Transmembrane; Thylakoid; Cyanelle.  
 FT DOMAIN 1 4 LUMENAL (POTENTIAL).  
 FT TRANSMEM 5 23 POTENTIAL.  
 FT DOMAIN 24 38 STROMAL (POTENTIAL).  
 SQ SEQUENCE 38 AA; 4240 MW; D11996A62BA74DFC CRC64;  
 Query Match 7.78; Score 49; DB 1; Length 38;  
 Best Local Similarity 34.48; Pred. No. 57;  
 Matches 11; Conservative 7; Mismatches 10; Indels 4; Gaps 1;  
 QY 9 SLRVAKVAYPLGLCVGVFIY----VAYIKWHR 36  
 DB 2 SMRLVVLLPLGIALGWAVYNGKLAIEQWR 33  
 RESULT 6  
 GAST HUMAN STANDARD; PRT; 101 AA.  
 AC P01350; P78463; P78464;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 23-OCT-1986 (Rel. 02, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Gastrin precursor.  
 GN GAS.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87219893; PubMed=3034736;  
 RA Kariya Y., Kato K., Hayashizaki Y., Himeno S., Tarui S., Matsubara K.;  
 RT "Expression of human gastrin gene in normal and gastrinoma tissues.";  
 RJ Gene 50:345-352(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84272693; PubMed=6087340;  
 RA Ito R., Sato K., Helmer T., Jay G., Agarwal K.L.;  
 RT "Structural analysis of the gene encoding human gastrin: the large  
 RT intron contains an Alu sequence.";  
 RJ Proc. Natl. Acad. Sci. U.S.A. 81:4662-4666(1984).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84169471; PubMed=6324077;  
 RA Kato K., Hayashizaki Y., Takahashi Y., Himeno S., Matsubara K.;  
 RT "Molecular cloning of the human gastrin gene.";  
 RJ Nucleic Acids Res. 11:8197-8203(1983).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=83221503; PubMed=6574456;  
 RA Boel E., Vuust J., Norris F., Norris K., Wind A., Rehfeld J.F.,  
 RA Marcker K.A.;  
 RT "Molecular cloning of human gastrin cDNA: evidence for evolution of  
 RT gastrin by gene duplication.";  
 RJ Proc. Natl. Acad. Sci. U.S.A. 80:2866-2869(1983).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84144842; PubMed=6322186;  
 RA Wiborg O., Berglund L., Boel E., Norris F., Norris K., Rehfeld J.F.,  
 RA Marcker K.A., Vuust J.;  
 RT "Structure of a human gastrin gene.";  
 RJ Proc. Natl. Acad. Sci. U.S.A. 81:1067-1069(1984).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84159488; PubMed=6689486;  
 RA Kato K., Himeno S., Takahashi Y., Wakabayashi T., Tarui S.,  
 RA Matsubara K.;

RT "Molecular cloning of human gastrin precursor cDNA.";  
 RL Gene 26:53-57(1983).  
 RN [7]  
 RP SEQUENCE OF 22-101, AND CHARACTERIZATION OF GASTRIN 71.  
 RC TISSUE=Antral mucosa;  
 RX MEDLINE=94333379; PubMed=8055952;  
 RA Rehfeld J.F., Johnsen A.H.;  
 RT "Identification of gastrin component I as gastrin-71. The largest  
 RT possible bioactive progastrin product.";  
 RJ Eur. J. Biochem. 223:765-773(1994).  
 RN [8]  
 RP SEQUENCE OF 76-92.  
 RX MEDLINE=67021327; PubMed=5921183;  
 RA Bentley P.H., Kenner G.W., Sheppard R.C.;  
 RT "Structures of human Gastrins I and II.";  
 RJ Nature 209:583-585(1966).  
 RN [9]  
 RP SEQUENCE OF 59-68.  
 RX MEDLINE=89273602; PubMed=2730647;  
 RA Higashimoto Y., Himeno S., Shinomura Y., Nagao K., Tamura T.,  
 RA Tarui S.;  
 RT "Purification and structural determination of urinary NH2-terminal  
 RT big gastrin fragments.";  
 RJ Biochem. Biophys. Res. Commun. 160:1364-1370(1989).  
 RN [10]  
 RP SEQUENCE OF 76-92.  
 RX MEDLINE=69298172; PubMed=5822140;  
 RA Gregory R.A., Tracy H.J., Agarwal K.L., Grossman M.I.;  
 RT "Amino acid constitution of two gastrins isolated from  
 RT Zollinger-Ellison tumour tissue.";  
 RJ Gut 10:603-608(1969).  
 RL CC  
 CC FUNCTION: Gastrin stimulates the stomach mucosa to produce and  
 CC secrete hydrochloric acid and the pancreas to secrete its  
 CC digestive enzymes. It also stimulates smooth muscle contraction  
 CC and increases blood circulation and water secretion in the stomach  
 CC and intestine.  
 CC  
 CC SUBCELLULAR LOCATION: Secreted.  
 CC  
 CC PTM: TWO DIFFERENT PROCESSING PATHWAYS PROBABLY EXIST IN ANTRAL  
 CC G-CELLS. IN THE DOMINANT PATHWAY PROGASTRIN IS CLEAVED AT THREE  
 CC SITES RESULTING IN TWO MAJOR BIOACTIVE GASTRINS, GASTRIN-34 AND  
 CC GASTRIN-17. IN THE PUTATIVE ALTERNATIVE PATHWAY, PROGASTRIN MAY  
 CC BE PROCESSED ONLY AT THE MOST C-TERMINAL DIABASIC SITE RESULTING  
 CC IN THE SYNTHESIS OF GASTRIN-71.  
 CC  
 CC SIMILARITY: Belongs to the gastrin/cholecystokinin family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X00183; CAA25005.1; -;  
 DR EMBL; X00183; CAA25006.1; -;  
 DR EMBL; X00183; CAA25007.1; -;  
 DR EMBL; V00511; CAA23769.1; -;  
 DR EMBL; M15958; AAA52520.1; -;  
 DR EMBL; K01254; AAB59533.1; -;  
 DR PIR; A91997; GMEHUB  
 DR Genew; HGNC:4164; GAS.  
 DR MIM; 137250; -;  
 DR GO; GO:0005180; F:peptide hormone; TAS.  
 DR InterPro; IPR001651; Gastrin.  
 DR Pfam; PF00918; Gastrin; 1.  
 DR SMART; SM00029; GASTRIN; 1.  
 DR PROSITE; PS00259; GASTRIN; 1.  
 KW Hormone; Cleavage on pair of basic residues; Amidation; Sulfation;  
 KW Signal; Phosphorylation; Pyrrolidone carboxylic acid.  
 FT SIGNAL 1 21  
 FT PEPTIDE 22 92  
 FT PEPTIDE 40 92  
 FT PEPTIDE 59 92  
 FT GASTRIN 71 (COMPONENT I).  
 FT GASTRIN 52 (POTENTIAL).  
 FT BIG GASTRIN (GASTRIN 34) (COMPONENT II).

```
FT PEPTIDE 76 92 GASTRIN (GASTRIN 17) (COMPONENT III).
FT MOD_RES 59 59 PYROLIDONE CARBOXYLIC ACID.
FT MOD_RES 76 76 PYROLIDONE CARBOXYLIC ACID.
FT MOD_RES 87 87 SULFATION (33%).
FT MOD_RES 92 92 AMIDATION (G-93 PROVIDE AMIDE GROUP).
FT MOD_RES 96 96 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 101 AA; 11394 MW; A03C847FCFE721C CRC64;

Query Match 7.7%; Score 49; DB 1; Length 101;
Best Local Similarity 35.3%; Pred. No. 1.7e+02;
Matches 18; Conservative 3; Mismatches 12; Indels 18; Gaps 3;

Oy 21 LCVGVFIYVAYIKMHRATATQAFSITRAAPGARW---GQAAHSPLGTAD 68
Db 4 LCVVVLIF-----ALALAASE-----ASMKRSQGPAPLGTGAN 39

RESULT 7
VE4_BPVA STANDARD; PRT; 107 AA.
AC P08347;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-AUG-1988 (Rel. 08, Last annotation update)
DE Probable B4 protein.
OS Bovine Papillomavirus type 4.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
CX NCBI_TaxID=10562;
RN 11
RP SEQUENCE FROM N.A.
RA MEDLINE=87282264; PubMed=3039043;
RX Patel K.R., Smith K.T., Campo M.S.;
RT "The nucleotide sequence and genome organization of bovine
RT Papillomavirus type 4";
RT J. Gen. Virol. 68:2117-2128(1987).

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X05817; NOT ANNOTATED_CDS.
DR PIR; G27129; W4WLB4.
KW Early protein.
SQ SEQUENCE 107 AA; 11846 MW; 1CC2A84EF30B4C2 CRC64;

Query Match 7.6%; Score 48.5; DB 1; Length 107;
Best Local Similarity 28.1%; Pred. No. 2e+02;
Matches 16; Conservative 10; Mismatches 20; Indels 11; Gaps 3;

Oy 13 AKVAVPLCLGCVFIYVAY---IKMHRATATQAFSITRAAPGARWQQAHSPLGT 65
Db 48 AEGAVVPAIATVITITVEGALQWTRHATE-----SRTPGRK-TKGHPRTGT 97

RESULT 8
GDIR_CAVPO STANDARD; PRT; 111 AA.
AC P80237;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Rho GDI-disassociation inhibitor 1 (Rho GDI 1) (Rho-GDI alpha)
DE (Fragments).
GN ARHGDI1.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Heteriognathi; Caviidae; Cavia.
CX NCBI_TaxID=10141;
```

```
RN [1]
RP STRAIN=Hartley; TISSUE=Macrophage;
RC MEDLINE=94039069; PubMed=8223583;
RX Pick E., Gorzalcany Y., Engel S.;
RA "Role of the rac1 p21-GDP-dissociation inhibitor for rho heterodimer
RT in the activation of the superoxide-forming NADPH oxidase of
RT macrophages.";
RT Eur. J. Biochem. 217:441-455(1993).
CC -1- FUNCTION: Regulates the GDP/GTP exchange reaction of the Rho
CC proteins by inhibiting the dissociation of GDP from them, and the
CC subsequent binding of GTP to them.
CC -1- FUNCTION: Rac1 p21/rho GDI heterodimer is the active component of
CC the cytosolic factor sigma 1, which is involved in stimulation of
CC the NADPH oxidase activity in macrophages.
CC -1- SUBUNIT: FORMS A HETERODIMER WITH P21/RAC-1.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- PTM: The N-terminus is blocked.
CC -1- SIMILARITY: Belongs to the Rho GDI family.
DR HSSP; P19803; IGDF.
DR InterPro; IPR007110; IG-1-like.
DR InterPro; IPR000406; Rho_GDI.
DR PRINTS; PR00492; RHODI.
KW GTPase activation.
FT NON_TER 1
FT NON_CONS 40 41
FT NON_CONS 46 47
FT NON_CONS 86 87
FT UNSURE 98 98
FT UNSURE 100 100
FT NON_TER 111
SQ SEQUENCE 111 AA; 12496 MW; 0FCE35BEA40FE951 CRC64;

Query Match 7.6%; Score 48.5; DB 1; Length 111;
Best Local Similarity 26.8%; Pred. No. 2.1e+02;
Matches 15; Conservative 5; Mismatches 21; Indels 15; Gaps 2;

Oy 70 HEVFGIMFD---AGSGTGVHVFOFTRPREPTITHEFKVXKGLSYAD 120
Db 50 HTRYKGVKIDKTDYVGVGYPRAEYEFILTPWEAP-----KQMLAFSTD 95

RESULT 9
PSBE_CHLVU STANDARD; PRT; 83 AA.
ID PSBE_CHLVU
AC P56309;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome b559 alpha subunit (PSII reaction center subunit V).
GN PSBE.
OS Chlorella vulgaris.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
OC Chlorellaceae; Chlorella.
OX NCBI_TaxID=3077;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=IAM C-27 / Tamiya;
RX MEDLINE=97303241; PubMed=9159184;
RA Wakaegi T., Nagai T., Kapoor M., Sugita M., Ito M., Ito S.,
RA Tsudzuki J., Nakashima K., Tsudzuki T., Suzuki Y., Hamada A., Ohta T.,
RA Inamura A., Yoshinaga K., Sugita M.;
RT "Complete nucleotide sequence of the chloroplast genome from the
RT green alga Chlorella vulgaris: the existence of genes possibly
RT involved in chloroplast division.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:5967-5972(1997).
CC -1- FUNCTION: This b-type cytochrome is tightly associated with the
CC reaction center of photosystem II and possibly is part of the
CC water-oxidation complex.
CC -1- SUBUNIT: Heterodimer of an alpha subunit and a beta subunit.
CC -1- SIMILARITY: Belongs to the psbE / psbF family.
CC -----
```





```
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitz E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grympey B.,
RA Welfens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leinacher H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler H., Warbut R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Laureau V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurt S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garçon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880 (2002).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC -----
CC EMBL; AL032824; CAB37425.1; -
CC PIR; T39132; S62526.
CC GeneDB Spombe; SPCC584.12; -.
CC Hypothetical protein.
CC SO SEQUENCE 115 AA; 13701 MW; 289D50B9A5CA13E CRC64;
QY Query Match 7.4%; Score 47.5; DB 1; Length 115;
Best Local Similarity 63.2%; Pred. No. 2.8e+02;
Matches 12; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
DB 93 TRPPEPTLTHTETPKAVK 111
93 TRPPEPTLTHTETPKAVK 24
DB 7 TRPPEPTLTHTETPKAVK 24
QY 93 TRPPEPTLTHTETPKAVK 111
YAHN ECOLI STANDARD; PRT; 106 AA.
ID YAHN ECOLI
AC P75650.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yahn.
DE YAHN OR B03322.
GN Escherichia coli.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OC NCBI_TaxID=562;
ON NCB1
RX SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirtpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474 (1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
CC -----
CC EMBL; AB000139; AAC73425.1; -
CC PIR; B64759; B64759.
CC Ecogene; EG13592; Yahn.
CC Hypothetical protein; Complete proteome.
CC SO SEQUENCE 106 AA; 10634 MW; 2260B0C42954AF71 CRC64;
QY Query Match 7.3%; Score 47; DB 1; Length 106;
Best Local Similarity 24.7%; Pred. No. 2.9e+02;
Matches 20; Conservative 13; Mismatches 32; Indels 16; Gaps 4;
DB 37 ATNQAFFSI-----TRAPARWGQQAHSPLGTADGHEFYGVIMPDAGSTGRVHVFQF 92
18 SVAHQAFTVTAAGCGVNLISLRMAQES---VGLISVAHQAF-----ATTAGCGVDALSG 68
QY 93 TRPPEPTLTHTETPKAV 110
69 LRVARBSVGLISVAHQAFVTI 89
DB 69 LRVARBSVGLISVAHQAFVTI 89
QY 93 TRPPEPTLTHTETPKAV 110
NU3M_CYPCA STANDARD; PRT; 116 AA.
ID NU3M_CYPCA
AC P24974;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 3 (EC 1.6.5.3).
GN MTND3 OR ND3.
OS Cyprinus carpio (Common carp).
OS Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Cyprini.
OC NCBI_TaxID=7962;
ON NCB1
RX SEQUENCE FROM N.A.
RC MEDLINE=94223691; PubMed=8169959;
RA Chang Y.S., Huang F.L., Lo T.B.;
RT "The complete nucleotide sequence and gene organization of carp
RL J. Mol. Evol. 38:138-155 (1994).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SIMILARITY: Belongs to the complex I subunit 3 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X61010; CAA4334.1; -
CC PIR; S36004; S36004.
CC InterPro; IPR000440; Oxidored_g4.
CC Pfam; PF00507; oxidored_g4; 1.
CC Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
CC SO SEQUENCE 116 AA; 13071 MW; DF5E99FD41DFEBA CRC64;
QY Query Match 7.3%; Score 47; DB 1; Length 116;
Best Local Similarity 63.6%; Pred. No. 3.2e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
DB 55 WGOQAHSPLGT 65
78 WGOQAHSPLGT 88
QY 55 WGOQAHSPLGT 65
WGOQAHSPLGT 88
DB 78 WGOQAHSPLGT 88
QY 78 WGOQAHSPLGT 88
NU3M_CYPCA STANDARD; PRT; 97 AA.
ID NU3M_CYPCA
AC P05678;
QY 97 WGOQAHSPLGT 65
WGOQAHSPLGT 88
DB 97 WGOQAHSPLGT 65
WGOQAHSPLGT 88
```

Search completed: June 8, 2004, 10:00:48  
Job time : 20 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 09:57:40 ; Search time 39 Seconds  
(without alignments)  
570.825 Million cell updates/sec

Title: US-09-905-589a-2\_copy\_1\_120  
Perfect score: 640  
Sequence: 1 MKRISNHSGLRVAKVAVPLD.....TLTHTFKAVKQLSAVDD 120

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 231205

Minimum DB seg length: 0  
Maximum DB seg length: 120

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	75.5	11.8	119	16	Q7UNK7 rhodospirill
2	69.5	10.9	64	12	Q8QJL2 mamestra co
3	60.5	9.5	105	2	Q83ZP7 escherichia
4	60	9.4	106	3	Q86ZEL phoma sp. c
5	60	9.4	114	16	Q91JQ3 streptomyce
6	59	9.2	102	8	Q9TEV4 opsanus bet
7	58.5	9.1	102	8	Q9TJH8 opsanus bet
8	58.5	9.1	102	8	Q9TEV3 opsanus par
9	58.5	9.1	102	8	Q9TEV8 porichnys
10	58	9.1	99	5	Q9STU9 opsanus tau
11	57.5	9.0	102	8	Q9TEV1 opsanus tau
12	57.5	9.0	102	8	Q9TEV2 opsanus tau
13	57.5	9.0	102	8	Q9B272 rhodospirill
14	57.5	9.0	105	8	Q9B272 rhodospirill
15	57.5	9.0	120	8	Q9T7C8 trinomys pa
16	57	8.9	70	16	Q8XWKL talstonia s

17	57	8.9	108	16	Q8NPP8	Q8NPP8 corynebacte
18	56.5	8.8	91	16	Q8ZUV9	Q8ZUV9 nitrosomona
19	56.5	8.8	101	8	Q20821	Q20821 lerista bou
20	56.5	8.8	102	8	Q20818	Q20818 lerista bou
21	56.5	8.8	104	8	Q20820	Q20820 lerista bou
22	56.5	8.8	105	8	Q9T7B8	Q9T7B8 trinomys ih
23	56.5	8.8	111	8	Q9T7B6	Q9T7B6 trinomys ih
24	56.5	8.8	118	8	Q9T7B7	Q9T7B7 trinomys ih
25	56	8.8	105	8	Q9B268	Q9B268 pachyactyl
26	56	8.8	115	17	Q9Y929	Q9Y929 aeropyrum p
27	55.5	8.7	100	8	Q47715	Q47715 cephalopnus
28	55.5	8.7	111	8	Q9MEJ5	Q9MEJ5 pronolagus
29	55	8.6	96	8	Q9B430	Q9B430 hemidactylu
30	54.5	8.5	73	16	Q9RDR5	Q9RDR5 streptomyce
31	54.5	8.5	85	16	Q7U6D9	Q7U6D9 synecococc
32	54.5	8.5	105	8	Q8WD29	Q8WD29 xantusia he
33	54.5	8.5	108	8	Q9T7D1	Q9T7D1 trinomys se
34	54.5	8.5	111	8	Q9MEJ9	Q9MEJ9 pronolagus
35	54.5	8.5	119	8	Q9MEK0	Q9MEK0 propithecus
36	54.5	8.5	119	8	Q9MEK8	Q9MEK8 lepitemur s
37	54	8.4	101	10	Q7XUB5	Q7XUB5 oryza sativ
38	54	8.4	119	16	Q9L220	Q9L220 streptomyce
39	53.5	8.4	67	16	Q87R81	Q87R81 vibrio para
40	53.5	8.4	96	8	Q86050	Q86050 tupia glis
41	53.5	8.4	98	8	Q85Q60	Q85Q60 hippidion s
42	53.5	8.4	104	8	Q9T7D5	Q9T7D5 trinomys se
43	53.5	8.4	111	8	Q9MEJ8	Q9MEJ8 pronolagus
44	53.5	8.4	113	16	Q8ZBW0	Q8ZBW0 yersinia pe
45	53.5	8.4	116	8	Q9T7D6	Q9T7D6 trinomys se

## ALIGNMENTS

## RESULT 1

Q7UNK7 PRELIMINARY; PRT; 119 AA.

AC Q7UNK7; 01-OCT-2003 (TREMBLrel. 25, Created)  
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
DE Hypothetical protein.  
GN R87531  
OS Rhodospirillum rubrum.  
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;  
OC Planctomycetaceae; Pirellula.  
OC NCBI\_TaxId=117;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=; PubMed=12835416;  
RX MEDLINE=22735913; PubMed=12835416;  
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,  
RA Ludwig W., Gade D., Beck A., Borzym K., Heilmann K., Rabus R.,  
RA Schlesner H., Amann R., Reinhardt R.;  
RT "Complete genome sequence of the marine planctomycete Pirellula sp.  
RT strain 1."  
RT Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).  
RU EMBL; BX294146; CAD75412.1;--  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 119 AA; 13173 MW; 98BAE8D8C6FFB8 CRC64;

Query Match 11.8%; Score 75.5; DB 16; Length 119;  
Best Local Similarity 30.3%; Pred. No. 1;  
Matches 36; Conservative 11; Mismatches 49; Indels 23; Gaps 8;

QY 6 NFGSLRVAKVAVPLDLCGVETIYVAYIKMRATATQAF-ESITRAPG---ARMQQAHS 61  
DB 4 NSGDASAKRRACPOHSGTGLIRVKSTIS---GLATIGDFCTVNAKGLVSRQVPHR 60  
QY 62 PLG-TAADQHE-VFYGIMEDA-----GSTGT---RVAVFOFTRPP-----RETPTLTH 104  
DB 61 PLGSLRREHQRVFLSWVIDCTHRTGRSGFGTTHRTITGLKFRKQGTITSCRKIPKPLH 119

```

RESULT 2
Q8QLI2 PRELIMINARY; PRT; 64 AA.
AC Q8QLI2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mamestra configurata nucleopolyhedrovirus.
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=191492;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=90/2;
RX MEDLINE=97163493; PubMed=9010313;
RA Li S., Erlandson M., Moody D., Gillett C.;
RT "A physical map of the Mamestra configurata nucleopolyhedrovirus
RT genome and sequence analysis of the polyhedrin gene.";
RL J. Gen. Virol. 78:265-271(1997).
RN [2];
RP SEQUENCE FROM N.A.
RC STRAIN=90/2;
RX MEDLINE=21884635; PubMed=11886270;
RA Li Q., Donly C., Li L., Willis L.G., Theilmann D.A., Erlandson M.;
RT "Sequence and Organization of the Mamestra configurata
RT Nucleopolyhedrovirus Genome.";
RL Virology 294:106-121(2002).
RN [3];
RP SEQUENCE FROM N.A.
RC STRAIN=90/2;
RA Li Q., Donly C., Li L., Willis L.G., Theilmann D.A., Erlandson M.A.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; U59461; AAM09160.1; -.
KW Hypothetical protein.
SQ SEQUENCE 64 AA; 7844 MW; 472234944556505A CRC64;

Query Match 10.9%; Score 69.5; DB 12; Length 64;
Best Local Similarity 33.3%; Pred. No. 2.3;
Matches 14; Conservative 10; Mismatches 11; Indels 7; Gaps 2;

QY 2 RKISNHSRLRVAKVPLGLCV-----GVFIYVAVIKWHRA 37
DB 14 RVTNHKHVKNKWRP-FICIIHPQCSWYVQYIRWHS 54

RESULT 3
Q83ZP7 PRELIMINARY; PRT; 105 AA.
AC Q83ZP7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative small multidrug resistance protein.
GN SMR.
OS Escherichia coli.
OG Plasmid PAK33.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1];
RP SEQUENCE FROM N.A.
RA Vourli S., Tzouvelekas L.S., Tzelepi E., Lebesse E., Legakis N.J.,
RA Miriagou V.;
RT "Integron In11, a novel class I integron from Escherichia coli, which
RT is bounded by IS26 insertion elements.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY260546; AAP22975.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR InterPro; IPR000390; Smr.
DR Pfam; PF00893; Multi_Drug_Res; 1.

KW Plasmid.
SQ SEQUENCE 105 AA; 11210 MW; 30B57F9D4C73884 CRC64;

Query Match 9.5%; Score 60.5; DB 2; Length 105;
Best Local Similarity 25.0%; Pred. No. 45;
Matches 18; Conservative 12; Mismatches 29; Indels 13; Gaps 3;

QY 14 KVAYPLGLCVGFIYVAVIKWHRAATATQAFPSITRAAPGARWQQAAHSLPLGTADGHEVF 73
DB 13 EIGWPVGLKMA---QVPETRWSCGVIAVAFMAVSGF---LLLAQRHPIFGTA----- 59
QY 74 YGIMFDAGSTGT 85
DB 60 YAVWTGIGAGT 71

RESULT 4
Q86ZE1 PRELIMINARY; PRT; 106 AA.
AC Q86ZE1;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE C-methyl transferase (fragment).
OS Phoma sp. C2932.
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Phoma.
OX NCBI_TaxID=86977;
RN [1];
RP SEQUENCE FROM N.A.
RX MEDLINE=21150635; PubMed=11251290;
RA Nicholson T.P., Rudd B.A.M., Dawson M., Lazarus C.M., Simpson T.J.,
RA Cox R.J.;
RT "Design and utility of oligonucleotide gene probes for fungal
RT polyketide synthases.";
RL Chem. Biol. 8:157-178(2001).
RN [2];
RP SEQUENCE FROM N.A.
RA Glod F., Cox R.J., Simpson T.J.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY217787; AAO62424.1; -.
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. .; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR001601; Methyltransf.
DR InterPro; IPR000051; SAM_bind.
KW Transferase.
FT NON TER 1
FT NON TER 106
SQ SEQUENCE 106 AA; 11440 MW; E9A1D80A0E2EE53E CRC64;

Query Match 9.4%; Score 60; DB 3; Length 106;
Best Local Similarity 36.8%; Pred. No. 52;
Matches 15; Conservative 4; Mismatches 22; Indels 0; Gaps 0;

QY 80 AGSTGTRVHVQFTRPRPTTTLTHTFKAVKPGLSAYADD 120
DB 4 AGTGGTYHVLERLRNPDGSSKATQVHFTDISPGFLAKAD 44

RESULT 5
Q9LIQ3 PRELIMINARY; PRT; 114 AA.
AC Q9LIQ3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein SCO6862.
GN SCO6862 OR SC7F9.14C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1];
RP SEQUENCE FROM N.A.

```

RC STRAIN-A3(2) / M145;  
RX MEDLINE-21996410; PubMed=12000953;  
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Godle A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
RA Rabinowitsch E., Rajadream M.A., Rutherford K., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
RA Warren T., Wierozorek A., Woodward J., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete Streptomyces  
RT coelicolor A3(2).";  
RL Nature 417:141-147(2002).  
DR EMBL; AL939129; CAB72367.1; -  
KM Hypothetical protein; Complete proteome.  
SQ SEQUENCE 114 AA; 12316 MW; 1BA5C6423312E16D CRC64;  
  
Query Match 9.4%; Score 60; DB 16; Length 114;  
Best Local Similarity 28.9%; Pred. No. 57;  
Matches 22; Conservative 11; Mismatches 31; Indels 12; Gaps 3;  
  
QY 33 KMRH--ATATGAFSITRAAPGARWGQQAHSPLGTADGHEVFQIMPDASTGRVAV 89  
DB 22 QMRQDTARTETATITLQSAKP-----ABEGMTALDVIALVQSGFSTASLVLTIA 74  
QY 90 PQFTRPPEPTLITHE 105  
DB 75 WRTRP--STPVVITIE 88  
  
RESULT 6  
QYQTEV4 PRELIMINARY; PRT; 102 AA.  
ID QYQTEV4  
AC QYQTEV4;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Cytochrome b (Fragment).  
GN Cytb OR Cytb-B.  
OS Oposanus beta.  
OG Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Paracanthopterygii; Batrachoididae; Oposanus.  
OX NCBI\_TaxID=95145;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Freshwater D.W., Kyhn-Hansen C., Sarver S.K., Walsh P.J.;  
RT "Phylogeny of Oposanus (Batrachoididae) inferred from multiple  
RT mitochondrial DNA sequences.";  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE  
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A  
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL  
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).  
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY  
CC BOUND TO THE PROTEIN (BY SIMILARITY).  
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,  
CC CYTOCHROME C1 AND THE RISKE PROTEIN (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.  
CC EMBL; AF165346; AAD48462.1; -  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0005746; C:mitochondrial electron transport chain; IEA.  
DR GO; GO:0005739; C:mitochondrion; IEA.  
DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR GO; GO:0006810; P:transport; IEA.  
DR InterPro; IPR005797; Cytb\_b6\_N.  
DR Pfam; PF00033; cytochrome\_b\_N; 1.  
DR PROSITE; PS00192; CYTOCHROME\_B\_HEME; 1.  
KM Electron transport; Heme; Respiratory chain; Transmembrane; Transport;  
KM Mitochondrion.

FT NON TER 1 1  
FT NON TER 102 102  
SQ SEQUENCE 102 AA; 11494 MW; B0B5D97908F301AA CRC64;  
  
Query Match 9.2%; Score 59; DB 8; Length 102;  
Best Local Similarity 28.8%; Pred. No. 64;  
Matches 23; Conservative 11; Mismatches 18; Indels 28; Gaps 5;  
  
QY 19 LGLCV-----GVFIYAYIKMRATATQAFSITRAAPGARWG---QQAHSPLGTADG 69  
DB 5 LGMCLIVQMTTGVIAMHYI-----PDATMAFPSTLAHSRDVNYGWLIRNTH-----ANG 54  
QY 70 HEVRY-----GIMFDA 80  
DB 55 ASVFEMCMYLIHGRIGYDS 74  
  
RESULT 7  
QYQTH8 PRELIMINARY; PRT; 102 AA.  
ID QYQTH8  
AC QYQTH8;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Cytochrome b (Fragment).  
GN Cytb OR Cytb-B.  
OS Oposanus beta.  
OG Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Paracanthopterygii; Batrachoididae; Oposanus.  
OX NCBI\_TaxID=95145;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Freshwater D.W., Kyhn-Hansen C., Sarver S.K., Walsh P.J.;  
RT "Phylogeny of Oposanus (Batrachoididae) inferred from multiple  
RT mitochondrial DNA sequences.";  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE  
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A  
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL  
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).  
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY  
CC BOUND TO THE PROTEIN (BY SIMILARITY).  
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,  
CC CYTOCHROME C1 AND THE RISKE PROTEIN (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.  
CC EMBL; AF165345; AAD48461.1; -  
DR EMBL; AF165344; AAD48460.1; -  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0005746; C:mitochondrial electron transport chain; IEA.  
DR GO; GO:0005739; C:mitochondrion; IEA.  
DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR GO; GO:0006810; P:transport; IEA.  
DR InterPro; IPR005797; Cytb\_b6\_N.  
DR Pfam; PF00033; cytochrome\_b\_N; 1.  
DR PROSITE; PS00192; CYTOCHROME\_B\_HEME; 1.  
KM Electron transport; Heme; Respiratory chain; Transmembrane; Transport;  
KM Mitochondrion.  
FT NON TER 1 1  
FT NON TER 102 102  
SQ SEQUENCE 102 AA; 11450 MW; B0B5D97908F651FF CRC64;  
  
Query Match 9.1%; Score 58.5; DB 8; Length 102;  
Best Local Similarity 30.4%; Pred. No. 73;  
Matches 21; Conservative 10; Mismatches 19; Indels 19; Gaps 4;  
  
QY 19 LGLCV-----GVFIYAYIKMRATATQAFSITRAAPGARWG---QQAHSPLGTADG 69  
DB 5 LGMCLIVQMTTGVIAMHYI-----PDATMAFPSTLAHSRDVNYGWLIRNTH-----ANG 54  
QY 70 HEVRYGIMF 78

```

OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neoceratonychia; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Batrachoididae; Forichthys.
NCBI_TaxID=45384;
[1] _SEQUENCE FROM N.A.
RP Freshwater D.W., Kyhn-Hansen C., Sarver S.K., Walsh P.J.;
RA "Phylogeny of Opsanus (Batrachoididae) inferred from multiple
RT mitochondrial DNA sequences.";
RL Submitted (Jul-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
DR EMBL; AF165352; RAD48468.1; -.
DR GO; GO:0013021; C:integral to membrane; IEA.
DR GO; GO:0005746; C:mitochondrial electron transport chain; IEA.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR005797; Cytb_b6_N.
DR Pfam; PF00033; cytochrome_b_1.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
KW Electron transport; Heme; Respiratory chain; Transmembrane; Transport;
KW Mitochondrion.
FT NON_TER 1
FT TER 102
FT SEQUENCE 102 AA; 11357 MW; 4E9DA2A8ADCE1B84 CRC64;
QY Query Match 9.1%; Score 58.5; DB 8; Length 102;
Db Best Local Similarity 30.4%; Pred. No. 73;
QY Matches 21; Conservative 9; Mismatches 20; Indels 19; Gaps 4;
QY 19 LGLCV-----GVFIYVVIKWEHATATCATQAFPSITRAAPGARWG---QQAHSPGLGTAADG 69
Db 5 LGLCLVTQMIVGFVIAHYI----PDTALFSSVAHLTRDYNVGWLLRNLIH-----ANG 54
QY 70 HEVFYGYIMF 78
Db :::::
QY 55 ASGIFFICMY 63
Db :::::

RESULT 10
Q95TD9 PRELIMINARY; PRT; 99 AA.
AC Q95TD9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DRC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE LD29295P.
DE CG14211.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
[1] _SEQUENCE FROM N.A.
RP STRAIN=Berkely;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise B., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Paclob J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY059453; AAIL3359.1; -.

```

DR FlyBase; FBgn031044; CG14211.  
SQ SEQUENCE 99 AA; 11674 MW; B8F1CA2E34448680 CRC64;

Query Match 9.1%; Score 58; DB 5; Length 99;  
Best Local Similarity 23.5%; Pred. No. 81;  
Matches 16; Conservative 15; Mismatches 25; Indels 12; Gaps 2;

QY 2 RKISNHSGLRVAKYAPLGLCVGIYVAYIKMR-----ATATQAFSTTRAPGARWG 56  
DB 16 RKTQHPSMTISKI-----YALYIHIIHLVYNNIVGQDTISVMSMERSPASWG 68  
QY 57 QOASPLG 64  
DB 69 LRAPSDIG 76

## RESULT 11

Q9TEVO PRELIMINARY; PRT; 95 AA.

AC O9TEV2; 01-MAY-2000 (TRENBLREL. 13, Created)  
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)  
DT 01-JUN-2003 (TRENBLREL. 24, Last annotation update)  
DE Cytochrome b (Fragment).  
GN CYTB OR CYT-B.  
OS Opsanus tau (Oyster toadfish).  
OC Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorphia; Paracanthopterygii; Batrachoididae; Opsanus.  
OX NCBI\_TaxID=8068;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Freshwater D.W., Kyhn-Hansen C., Sarver S.K., Walsh P.J.;  
RT "Phylogeny of Opsanus (Batrachoididae) inferred from multiple  
mitochondrial DNA sequences."  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE  
COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A  
RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL  
COUPLED TO ATP SYNTHESIS (BY SIMILARITY).  
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY  
BOUND TO THE PROTEIN (BY SIMILARITY).  
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,  
CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.  
DR EMBL; AF165350; AAD48466.1; -.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0005746; C:mitochondrial electron transport chain; IEA.  
DR GO; GO:0005739; C:mitochondrion; IEA.  
DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR GO; GO:0006810; P:transport; IEA.  
DR InterPro; IPR005797; Cytb b6 N.  
DR Pfam; PF00033; Cytochrome\_b\_N.1  
DR PROSITE; PS00192; CYTOCHROME\_B\_HEME; 1.  
KW Electron transport; Heme; Respiratory chain; Transmembrane; Transport;  
KM Mitochondrion.  
FT NON TER 1  
FT NON TER 95  
SQ SEQUENCE 95 AA; 10763 MW; 8A53F22539E51FB CRC64;

Query Match 9.0%; Score 57.5; DB 8; Length 95;  
Best Local Similarity 29.0%; Pred. No. 87;  
Matches 20; Conservative 11; Mismatches 19; Indels 19; Gaps 4;

QY 19 LGLCV-----GVFIYVAYIKMRATQAFSSITRAPGARWG---QOASPLGTADG 69  
DB 1 LGMCLIVQMAIGVFIAMRYI-----PDATMAPPSLAHSRDVNYGLNLNIH-----ANG 50  
QY 70 HEVFIYGV 78  
DB 51 ASIFMCMY 59

## RESULT 12

Q9TEV1 PRELIMINARY; PRT; 102 AA.

AC O9TEV1; 01-MAY-2000 (TRENBLREL. 13, Created)  
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)  
DT 01-JUN-2003 (TRENBLREL. 24, Last annotation update)  
DE Cytochrome b (Fragment).  
GN CYTB OR CYT-B.  
OS Opsanus tau (Oyster toadfish).  
OC Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorphia; Paracanthopterygii; Batrachoididae; Opsanus.  
OX NCBI\_TaxID=8068;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Freshwater D.W., Kyhn-Hansen C., Sarver S.K., Walsh P.J.;  
RT "Phylogeny of Opsanus (Batrachoididae) inferred from multiple  
mitochondrial DNA sequences."  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE  
COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A  
RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL  
COUPLED TO ATP SYNTHESIS (BY SIMILARITY).  
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY  
BOUND TO THE PROTEIN (BY SIMILARITY).  
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,  
CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.  
DR EMBL; AF165349; AAD48465.1; -.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0005746; C:mitochondrial electron transport chain; IEA.  
DR GO; GO:0005739; C:mitochondrion; IEA.  
DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR GO; GO:0006810; P:transport; IEA.  
DR InterPro; IPR005797; Cytb b6 N.  
DR Pfam; PF00033; Cytochrome\_b\_N.1  
DR PROSITE; PS00192; CYTOCHROME\_B\_HEME; 1.  
KW Electron transport; Heme; Respiratory chain; Transmembrane; Transport;  
KM Mitochondrion.  
FT NON TER 1  
FT NON TER 102  
SQ SEQUENCE 102 AA; 11430 MW; 0A00EF34A0ED687B CRC64;

Query Match 9.0%; Score 57.5; DB 8; Length 102;  
Best Local Similarity 29.0%; Pred. No. 95;  
Matches 20; Conservative 11; Mismatches 19; Indels 19; Gaps 4;

QY 19 LGLCV-----GVFIYVAYIKMRATQAFSSITRAPGARWG---QOASPLGTADG 69  
DB 5 LGMCLIVQMAIGVFIAMRYI-----PDATMAPPSLAHSRDVNYGLNLNIH-----ANG 54  
QY 70 HEVFIYGV 78  
DB 55 ASIFMCMY 63  
RESULT 13  
Q9TEV2 PRELIMINARY; PRT; 102 AA.  
AC O9TEV2; 01-MAY-2000 (TRENBLREL. 13, Created)  
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)  
DT 01-JUN-2003 (TRENBLREL. 24, Last annotation update)  
DE Cytochrome b (Fragment).  
GN CYTB OR CYT-B.  
OS Opsanus phobeton.  
OC Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 CC Acanthomorpha; Paracanthopterygii; Batrachoididae; Opsanus.  
 CC NCBI\_TaxID=101190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Freshwater D.W., Kynn-Hansen C., Sarver S.K., Walsh P.J.;  
 RT "Phylogeny of Opsanus (Batrachoididae) inferred from multiple  
 RL mitochondrial DNA sequences."  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE  
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A  
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL  
 CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).  
 CC -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY  
 CC BOUND TO THE PROTEIN (BY SIMILARITY).  
 CC -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,  
 CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.  
 DR EMBL: AF165348; AAD48464.1; --  
 DR GO: GO:0016021; C:integral to membrane; IEA.  
 DR GO: GO:0005746; C:mitochondrial electron transport chain; IEA.  
 DR GO: GO:0005739; C:mitochondrion; IEA.  
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO: GO:0006118; P:electron transport; IEA.  
 DR GO: GO:0006810; P:transport; IEA.  
 DR InterPro: IPR005797; Cytb\_b6 N.  
 DR Pfam: PF00033; cytochrome\_b\_N; 1.  
 DR PROSITE: PS00192; CYTOCHROME\_B\_HEME; 1.  
 KW Electron transport; Heme; Respiratory chain; Transmembrane; Transport;  
 KW Mitochondrion.  
 FT NON\_TER 1  
 FT NON\_TER 102 102  
 FT SEQUENCE 102 AA; 11411 MW; 6BD5D97908EF784D CRC64;  
 QY Query Match 9.0%; Score 57.5; DB 8; Length 102;  
 Best Local Similarity 29.0%; Pred. No. 95;  
 Matches 20; Conservative 11; Mismatches 19; Indels 19; Gaps 4;  
 QY 19 LGCLV-----GVFIYVAYIKWHRATATQAFSITRAAPGARWG---QQAHSPLGTADG 69  
 Db 5 LGMCLLVQMTTCGVFIAMHYI-----PDATWAPSLAHLSDVNYGMLLENIH-----ANG 54  
 QY 70 HEVFGIMF 78  
 Db 55 ASIFFCMY 63  
 RESULT 14  
 Q9B272 PRELIMINARY; PRT; 105 AA.  
 AC Q9B272;  
 DT 01-JUN-2001 (TREMELrel. 17, Created)  
 DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)  
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
 DE Cytochrome b (Fragment).  
 GN CYTB.  
 OS Rhothropus boultoni.  
 OG Mitochondrion.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Lepidodonta; Squamata; Scleroglossa; Gekkota; Gekkonidae; Rhothropus.  
 CC NCBI\_TaxID=152588;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Lamb T., Bauer A.M.;  
 RT "Mitochondrial phylogeny of Namib day geckos (Rhothropus) based on  
 RL cytochrome b and 16S rRNA sequences."  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE  
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A  
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL  
 CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).  
 CC -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY  
 CC BOUND TO THE PROTEIN (BY SIMILARITY).

CC -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,  
 CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.  
 DR EMBL: AY026922; AAK13033.1; --  
 DR GO: GO:0016021; C:integral to membrane; IEA.  
 DR GO: GO:0005746; C:mitochondrial electron transport chain; IEA.  
 DR GO: GO:0005739; C:mitochondrion; IEA.  
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO: GO:0006118; P:electron transport; IEA.  
 DR GO: GO:0006810; P:transport; IEA.  
 DR InterPro: IPR005797; Cytb\_b6 N.  
 DR Pfam: PF00033; cytochrome\_b\_N; 1.  
 DR PROSITE: PS00192; CYTOCHROME\_B\_HEME; 1.  
 KW Electron transport; Heme; Respiratory chain; Transmembrane; Transport;  
 KW Mitochondrion.  
 FT NON\_TER 1  
 FT NON\_TER 105 105  
 FT SEQUENCE 105 AA; 11701 MW; 16D0905FAL2AD39F CRC64;  
 QY Query Match 9.0%; Score 57.5; DB 8; Length 105;  
 Best Local Similarity 30.8%; Pred. No. 99;  
 Matches 24; Conservative 9; Mismatches 18; Indels 27; Gaps 5;  
 QY 6 NHGSLVAVKAVYPLGLCV-----GVFIYVAYIKWHRATATQAFSITRAAPGARWG--- 56  
 Db 28 NFGSL-----LGLCLLQIQTSTGLFLAMHY---SADATLAFNSIAHLCREVQFGMLL 75  
 QY 57 QQAHSPLGTAAADGHEVY 74  
 Db 76 RNLH-----ADGSMFF 87  
 RESULT 15  
 Q9T7C8 PRELIMINARY; PRT; 120 AA.  
 AC Q9T7C8;  
 DT 01-MAY-2000 (TREMELrel. 13, Created)  
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)  
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
 DE Cytochrome b (Fragment).  
 GN CYTB OR CYT B.  
 OS Trinomys paratus.  
 OG Mitochondrion.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Hystricognathi; Echimyidae; Trinomys.  
 CC NCBI\_TaxID=42827;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Lara M.C., Patton J.L.;  
 RT "Evolutionary diversification of spiny rats (genus Trinomys, Rodentia:  
 RL Echimyidae) in the Atlantic Forest of Brazil."  
 RL Zool. J. Linn. Soc. 0:0-0(2000).  
 CC -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE  
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A  
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL  
 CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).  
 CC -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY  
 CC BOUND TO THE PROTEIN (BY SIMILARITY).  
 CC -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,  
 CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.  
 DR EMBL: AF194294; AAF16096.1; --  
 DR GO: GO:0016021; C:integral to membrane; IEA.  
 DR GO: GO:0005746; C:mitochondrial electron transport chain; IEA.  
 DR GO: GO:0005739; C:mitochondrion; IEA.  
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO: GO:0006118; P:electron transport; IEA.  
 DR GO: GO:0006810; P:transport; IEA.  
 DR InterPro: IPR005797; Cytb\_b6 N.  
 DR Pfam: PF00033; cytochrome\_b\_N; 1.  
 DR PROSITE: PS00192; CYTOCHROME\_B\_HEME; 1.  
 KW Electron transport; Heme; Respiratory chain; Transmembrane; Transport;  
 KW Mitochondrion.



FT NON TER 120 120  
SQ SEQUENCE 120 AA; 13696 MW; 6826EA63B139162B CRC64;

Query Match 9.0%; Score 57.5; DB 8; Length 120;  
Best Local Similarity 26.8%; Pred. No. 1.2e+02;  
Matches 22; Conservative 13; Mismatches 20; Indels 27; Gaps 5;

QY 6 NHGSLRVAKVAYPPGLCV-----GVPIYVAYIKMHRATATQAFPSITRAAPGARWG--- 56  
DB 32 NPGSL-----LGVCLALQIITGLFLAMHY-----TADTTAFSSVTHICRDVNYGMLI 79  
QY 57 QQASPLGTADGHEVYGYMF 78  
DB 80 RYAH-----ANGASMPFIFLY 95

Search completed: June 8, 2004, 10:01:40  
Job time : 42 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: June 8, 2004, 09:51:40 ; Search time 54 Seconds  
(without alignments)  
627.883 Million cell updates/sec

Title: US-09-905-589a-2\_COPY\_1\_120  
Perfect score: 640  
Sequence: 1 MKRISNHSGLRVAKVAVPIG.....TTHETFAVPGLSAVYDD 120

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 948883

Minimum DB seq length: 0  
Maximum DB seq length: 120

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: A\_Geneseq\_290a04:\*  
2: geneseqp1380s:\*  
3: geneseqp1990s:\*  
4: geneseqp2000s:\*  
5: geneseqp2001s:\*  
6: geneseqp2002s:\*  
7: geneseqp2003as:\*  
8: geneseqp2003bs:\*  
9: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	414	64.7	105	5	ABP69389 Human pol
2	78	12.2	14	4	AA872245 Peptide u
3	72	11.2	95	4	AAU66349 Propionib
4	72	11.2	95	4	AAU64628 Propionib
5	72	11.2	95	6	ABM61147 Propionib
6	72	11.2	95	6	ABM62868 Propionib
7	63	9.8	92	5	ABG00152 Human sec
8	62	9.7	92	5	ABM78931 Tumour ne
9	61.5	9.6	90	4	AAU66429 Propionib
10	61.5	9.6	90	6	ABM62948 Propionib
11	60.5	9.5	100	7	ADD26584 BGS-19 im
12	58.5	9.1	114	4	AAU31171 Novel hum
13	58	9.1	82	4	ABM40819 Peptide #
14	58	9.1	82	4	AAU34585 Peptide #
15	58	9.1	82	4	AAU74471 Human bon
16	58	9.1	82	4	AAU61678 Human bra
17	58	9.1	82	5	ABG44353 Human pep
18	58	9.1	94	5	ABM78894 Tumour ne
19	57.5	9.0	112	5	ABM89458 Human pol
20	57	8.9	65	3	AAU36092 Zea may
21	57	8.9	72	4	AAU13641 Peptide #
22	57	8.9	72	4	AAU52571 Peptide #
23	57	8.9	72	4	AAU62608 Peptide #
24	57	8.9	72	4	ABM27422 Human pep
25	57	8.9	72	4	ABM18070 Protein #

26	57	8.9	72	4	AAU65779 Human bon
27	57	8.9	72	4	AAU53403 Human bra
28	57	8.9	72	4	ABG47424 Human liv
29	57	8.9	72	4	AAU01391 Peptide #
30	57	8.9	72	5	ABG35411 Human pep
31	57	8.9	85	4	AAU67744 Amilo aci
32	57	8.9	91	4	AAU85675 Human imm
33	57	8.9	92	5	ABP09987 Human ORF
34	57	8.9	108	4	AAU91689 C glutami
35	56.5	8.8	86	5	ABP59185 NADH dehy
36	56	8.8	94	5	ABM78867 Tumour ne
37	56	8.8	99	4	AAU59972 Propionib
38	56	8.8	99	6	ABM56491 Propionib
39	55.5	8.7	55	4	AAU04630 Human pol
40	55.5	8.7	58	3	AAU70926 Human CD3
41	55.5	8.7	58	3	AAU70903 Protein u
42	55.5	8.7	66	4	AAU65466 Propionib
43	55.5	8.7	66	6	ABM61985 Propionib
44	55.5	8.7	92	5	ABM78893 Tumour ne
45	55.5	8.7	106	5	ABP03812 Human ORF

## ALIGNMENTS

RESULT 1  
ABP69389  
ID ABP69389 standard; protein, 105 AA.  
XX  
AC ABP69389;  
XX  
DT 20-JAN-2003 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 1436.  
XX  
KW Human; genome mapping; gene therapy; food supplement; virus; fungus;  
cell-proliferative disorder; neurodegenerative disease; bacterial;  
Parkinson's disease; Alzheimer's disease; autoimmune disease;  
multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
arthritis; cytostatic; immunomodulator; neurologic; neuroprotective;  
antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;  
antiarthritic.  
XX  
KW Homo sapiens.  
OS  
XX  
PN WO200270539-A2.  
XX  
PD 12-SEP-2002.  
XX  
PF 05-MAR-2002; 2002WO-US005095.  
XX  
PR 05-MAR-2001; 2001US-00799451.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YJ, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
Wehman T, Wang J, Wang D, Drmanac RT;  
XX  
PI WPI; 2002-759812/82.  
XX  
DR N-PBDB; ABZ11606.  
XX  
DR New polynucleotides comprising sequences assembled from expressed  
PT sequence tags (ESTs), useful for treating cell-proliferative,  
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet  
PT or coagulation disorders.  
XX  
PS Claim 9; SEQ ID NO 1436; 1012bp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated polynucleotide (I) comprising a  
CC nucleotide sequence selected from any of 948 sequences (ABZ1119-  
CC ABZ12066) or their mature protein coding portion, active domain coding

CC protein or complementary sequences. The polynucleotides are useful for  
 CC identifying expressed genes or for physical mapping of human genome. The  
 CC encoded polypeptides (ABP68902-ABP69849) are useful as molecular weight  
 CC markers, as a food supplement, for generating antibodies, in medical  
 CC imaging, screening and diagnostic assays and for treating cell-  
 CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's  
 CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,  
 CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,  
 CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver  
 CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),  
 CC arthritis, etc. Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 105 AA;

Query Match 64.7%; Score 414; DB 5; Length 105;  
 Best Local Similarity 80.2%; Pred. No. 2.1e-40;  
 Matches 77; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY 1 MRKISNHSGLRVAKVAYPLGLCVGYVYVYIKWHRATATQAFSTRAAPGARWQQAH 60  
 DB 1 MRKIPNHGTLRWTKVAYPLGLCVGLFYVYIKWHRASAAQAFSTIAGASGARWTQAF 60

QY 61 SPLGTAADGHEVPYIGIMFDAGSTGTRVHVQFTRPP 96  
 DB 61 SPFGSAARGHEVPYIGIMFDAGSTGTRVHVQFARPP 96

RESULT 2  
 AAB72245  
 ID AAB72245 standard; peptide; 14 AA.  
 XX  
 AC AAB72245;  
 XX  
 DT 14-MAY-2001 (first entry)  
 XX  
 DE Peptide used to create anti human CD39-L2 antibodies.  
 XX  
 KW Human CD39-like protein; apyrase; NDPase; platelet function inhibitor;  
 KW myocardial infarction; cerebral ischaemia; angina; arterial thrombosis;  
 KW cerebral artery thrombosis; platelet aggregation; inflammation;  
 KW apoptosis; autoimmune disorder; neurological disorder; antibody;  
 KW Alzheimer's disease; Parkinson's disease; cancer; CD39-L2.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200110205-A1.  
 XX  
 XX 15-FEB-2001.  
 XX  
 XX 09-AUG-2000; 2000WO-US021790.  
 XX  
 XX 09-AUG-1999; 99US-00370265.  
 XX 11-JAN-2000; 2000US-00481238.  
 XX 25-APR-2000; 2000US-00557800.  
 XX 26-MAY-2000; 2000US-00583231.  
 XX 30-JUN-2000; 2000US-00608285.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 XX  
 XX Ford J, Mulero JJ, Yeung G;  
 XX WPI; 2001-147489/15.  
 XX  
 XX Polynucleotides encoding human CD39-like polypeptides, with apyrase  
 XX and/or NDPase activity, which are useful in the treatment of pathological  
 XX conditions caused by thrombosis (e.g. myocardial infarction) and  
 XX inflammatory disorders.  
 XX  
 XX Example 18; Page 102; 203pp; English.  
 XX  
 XX This invention relates to polynucleotides encoding human CD39-like

CC polypeptides with apyrase and/or NDPase activity. The polypeptides having  
 CC APDase, including NDPase, activity are useful for inhibiting platelet  
 CC function and can therefore be used in the prophylaxis or treatment of  
 CC pathological conditions caused by or involving thrombosis or excessive  
 CC coagulation or excessive platelet aggregation, such as myocardial  
 CC infarction, cerebral ischaemia, angina, arterial thrombosis, cerebral  
 CC artery thrombosis or intracardiac thrombosis, and conditions associated  
 CC with venous thrombosis. CD39-L4 and CD39-L2 polypeptides are useful in  
 CC modulating disease states (including platelet aggregation, inflammation  
 CC and apoptosis) associated with ADP or other purinergic signalling by  
 CC reducing the levels of NDPs. The polypeptides are also useful for  
 CC prophylaxis or treatment of inflammation related disorders, such as  
 CC disorders involving sepsis or systemic inflammatory response syndrome or  
 CC SIRS (and associated conditions such as fever, tachycardia, tachypnea,  
 CC cytokine overstimulation); autoimmune disorders such as thrombosis,  
 CC atherosclerosis, acute pancreatitis, dermatitis, including psoriasis;  
 CC cirrhosis, reperfusion injury, asthma, multiple sclerosis, arthritis;  
 CC neurological disorders including neurodegenerative diseases, epilepsy,  
 CC depression, Alzheimer's disease, Parkinson's disease, Huntington's  
 CC disease, and amyotrophic lateral sclerosis; and cancer. The present  
 CC sequence represents a peptide used to create antibodies directed to human  
 CC CD39 like protein CD39-L2  
 XX  
 SQ Sequence 14 AA;

Query Match 12.2%; Score 78; DB 4; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 0.033;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 TRPPRETPTLTTHET 106  
 DB 1 TRPPRETPTLTTHET 14  
 |||||  
 |||||

RESULT 3  
 AAU66349  
 ID AAU66349 standard; protein; 95 AA.  
 XX  
 AC AAU66349;  
 XX  
 DT 27-FEB-2002 (first entry)  
 XX  
 DE Propionibacterium acnes immunogenic protein #27245.  
 XX  
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW urethritis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 XX WO200181581-A2.  
 XX  
 XX 01-NOV-2001.  
 XX  
 XX 20-APR-2001; 2001WO-US012865.  
 XX  
 XX 21-APR-2000; 2000US-0199047P.  
 XX 02-JUN-2000; 2000US-0208841P.  
 XX 07-JUL-2000; 2000US-0216747P.  
 XX  
 XX (CORI-) CORIXA CORP.  
 XX  
 XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 XX L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 XX  
 XX WPI; 2001-616774/71.  
 XX N-PSDB; AAS99729.  
 XX  
 XX Propionibacterium acnes polypeptides and nucleic acids useful for  
 XX vaccinating against and diagnosing infections, especially useful for  
 XX treating acne vulgaris.

PS Example 1; SEQ ID NO 27544; 1069pp; English.

CC Sequences AAU39105-AAU68017 represent *Propionibacterium acnes* immunogenic polypeptides. The proteins and their associated DNA sequences are used in the treatment, prevention and diagnosis of medical conditions caused by *P. acnes*. The disorders include SAPHO syndrome (synovitis, acne, pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis. *P. acnes* is also involved in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the presence or absence of *P. acnes* in a patient comprises contacting a sample with a binding agent that binds to the proteins of the invention and determining the amount of bound protein in the sample. The polypeptides may be used as antigens in the production of antibodies specific for *P. acnes* proteins. These antibodies can be used to downregulate expression and activity of *P. acnes* polypeptides and therefore treat *P. acnes* infections. The antibodies may also be used as diagnostic agents for determining *P. acnes* presence, for example, by enzyme linked immunosorbent assay (ELISA). Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

CC Sequence 95 AA;

SO

Query Match 11.2%; Score 72; DB 4; Length 95;  
Best Local Similarity 34.8%; Pred. No. 2;  
Matches 23; Conservative 3; Mismatches 14; Indels 26; Gaps 4;

QY 36 RATAQAFFSI-TRAPGARWGQAHSPLGTADGHEVFYIMPDAGSTGRVHFQPTR 94  
DB 55 RALGTAAPRLDRLRDPGLRW--DAHSG-GTAPDSHRI----- 89  
QY 95 PPRETP 100  
DB 90 PPRESP 95

RESULT 4  
AAU64628  
ID AAU64628 standard; protein; 95 AA.

AC AAU64628;  
DT 27-FEB-2002 (first entry)

DE *Propionibacterium acnes* immunogenic protein #25524.

XX SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;  
KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KM dermatological; osteopathic; neuroprotectant.

XX *Propionibacterium acnes*.

OS

PN WO200181581-A2.

PD 01-NOV-2001.

PF 20-APR-2001; 2001WO-US012865.

XX 21-APR-2000; 2000US-0199047P.  
PR 02-JUN-2000; 2000US-020841P.  
PR 07-JUL-2000; 2000US-0216747P.

XX (CORI-) CORIXA CORP.

PA Skelky YAW, Persing DH, Mitcham JL, Wang SG, Bhacia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX WPI; 2001-616774/71.  
DR N-PSDB; AAS59646.

PT *Propionibacterium acnes* polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris.

PS Example 1; SEQ ID NO 25823; 1069pp; English.

CC Sequences AAU39105-AAU68017 represent *Propionibacterium acnes* immunogenic polypeptides. The proteins and their associated DNA sequences are used in the treatment, prevention and diagnosis of medical conditions caused by *P. acnes*. The disorders include SAPHO syndrome (synovitis, acne, pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis. *P. acnes* is also involved in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the presence or absence of *P. acnes* in a patient comprises contacting a sample with a binding agent that binds to the proteins of the invention and determining the amount of bound protein in the sample. The polypeptides may be used as antigens in the production of antibodies specific for *P. acnes* proteins. These antibodies can be used to downregulate expression and activity of *P. acnes* polypeptides and therefore treat *P. acnes* infections. The antibodies may also be used as diagnostic agents for determining *P. acnes* presence, for example, by enzyme linked immunosorbent assay (ELISA). Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

CC Sequence 95 AA;

SO

Query Match 11.2%; Score 72; DB 4; Length 95;  
Best Local Similarity 34.8%; Pred. No. 2;  
Matches 23; Conservative 3; Mismatches 14; Indels 26; Gaps 4;

QY 36 RATAQAFFSI-TRAPGARWGQAHSPLGTADGHEVFYIMPDAGSTGRVHFQPTR 94  
DB 55 RALGTAAPRLDRLRDPGLRW--DAHSG-GTAPDSHRI----- 89  
QY 95 PPRETP 100  
DB 90 PPRESP 95

RESULT 5  
ABM61147  
ID ABM61147 standard; protein; 95 AA.

AC ABM61147;  
DT 20-OCT-2003 (first entry)

DE *Propionibacterium acnes* predicted ORF encoded polypeptide #25823.

XX *Acne vulgaris*; antiseborrheic; dermatological; antibacterial;  
KM immunostimulant; immune response; vaccine.

XX *Propionibacterium acnes*.

OS

PN WO2003033515-A1.

PD 24-APR-2003.

XX 11-OCT-2002; 2002WO-US032727.  
PF 15-OCT-2001; 2001US-00978825.

XX (CORI-) CORIXA CORP.

PA Mitcham JL, Skelky YAW, Persing DH, Bhacia A, Maisonneuve JL;  
PI Zhang Y, Wang S, Lodes MJ, Benson DR, Jones R, Carter D;  
XX Barth B, Vallieue-Douglas J;  
DR WPI; 2003-381789/36.  
DR N-PSDB; ACF64575.

XX New Propionibacterium acnes polypeptides and polynucleotides encoding the  
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
PT or for stimulating an immune response specific for a P. acnes protein.  
XX  
PS Example 1; SEQ ID NO 25923; 1481pp; English.  
XX  
CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
CC encoding a Propionibacterium acnes protein. The invention also relates to  
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
CC immunogenic fragments of P. acnes polypeptides. The invention  
CC additionally encompasses expression vectors and host cells comprising a  
CC polynucleotide of the invention; antibodies against polypeptides of the  
CC invention; fusion proteins comprising a polypeptide of the invention; a  
CC method for stimulating an immune response specific for a P. acnes  
CC polypeptide and an isolated T cell population comprising T cells prepared  
CC via this method; a vaccine composition (comprising P. acnes polypeptides,  
CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
CC antigen-presenting cells that express the polypeptide); a method and kit  
CC for detecting or determining the presence or absence of P. acnes in a  
CC patient; and a method for inhibiting the development of P. acnes in a  
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
CC proteins, T cell populations or antigen-presenting cells that express the  
CC polypeptides are useful for diagnosing, preventing or treating acne  
CC vulgaris, or for stimulating an immune response specific for a P. acnes  
CC protein. The polynucleotides can also be used as probes or primers for  
CC nucleic acid hybridisation. The vaccine composition is useful for the  
CC stimulation of an immune response against P. acnes, or for treating acne,  
CC and the kit is useful for performing a diagnostic assay. The present  
CC sequence represents a polypeptide predicted to be encoded by an ORF (open  
CC reading frame) contained within the P. acnes polynucleotides of the  
CC invention. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 95 AA;  
XX  
Query Match 11.2%; Score 72; DB 6; Length 95;  
Best Local Similarity 34.8%; Pred. No. 2;  
Matches 23; Conservative 3; Mismatches 14; Indels 26; Gaps 4;  
Qy 36 RATATQAFPSI-TRAAPGARWGOQAHSPGLTAADGHEVFGIMFDAGSTGTRVHVQFTR 94  
Db 55 RALGTAAPFLDLRLDRPGLRW--DAHSG-GTAPDSHRI----- 89  
Qy 95 PPRETP 100  
Db 90 PPRES 95  
RESULT 6  
ABM62868  
XX ABM62868 standard; protein; 95 AA.  
XX  
AC ABM62868;  
XX  
XX 20-OCT-2003 (first entry)  
XX Propionibacterium acnes predicted ORF-encoded polypeptide #27544.  
XX  
XX Acne vulgaris; antiseborrheic; dermatological; antibacterial;  
XX immunostimulant; immune response; vaccine.  
XX  
XX Propionibacterium acnes.  
XX  
XX WO2003033515-A1.  
XX  
XX 24-APR-2003.  
XX  
XX 11-OCT-2002; 2002WO-US032727.  
XX  
XX 15-OCT-2001; 2001US-00978825.  
XX  
XX

PA (CORI-) CORIXA CORP.  
XX  
XX Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;  
PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
PI Barth B, Valliave-Douglass J;  
XX  
XX WPI; 2003-381789/36.  
DR N-PSDB; ACF64658.  
XX  
XX New Propionibacterium acnes polypeptides and polynucleotides encoding the  
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
PT or for stimulating an immune response specific for a P. acnes protein.  
XX  
PS Example 1; SEQ ID NO 27544; 1481pp; English.  
XX  
CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
CC encoding a Propionibacterium acnes protein. The invention also relates to  
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
CC immunogenic fragments of P. acnes polypeptides. The invention  
CC additionally encompasses expression vectors and host cells comprising a  
CC polynucleotide of the invention; antibodies against polypeptides of the  
CC invention; fusion proteins comprising a polypeptide of the invention; a  
CC method for stimulating an immune response specific for a P. acnes  
CC polypeptide and an isolated T cell population comprising T cells prepared  
CC via this method; a vaccine composition (comprising P. acnes polypeptides,  
CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
CC antigen-presenting cells that express the polypeptide); a method and kit  
CC for detecting or determining the presence or absence of P. acnes in a  
CC patient; and a method for inhibiting the development of P. acnes in a  
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
CC proteins, T cell populations or antigen-presenting cells that express the  
CC polypeptides are useful for diagnosing, preventing or treating acne  
CC vulgaris, or for stimulating an immune response specific for a P. acnes  
CC protein. The polynucleotides can also be used as probes or primers for  
CC nucleic acid hybridisation. The vaccine composition is useful for the  
CC stimulation of an immune response against P. acnes, or for treating acne,  
CC and the kit is useful for performing a diagnostic assay. The present  
CC sequence represents a polypeptide predicted to be encoded by an ORF (open  
CC reading frame) contained within the P. acnes polynucleotides of the  
CC invention. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 95 AA;  
XX  
Query Match 11.2%; Score 72; DB 6; Length 95;  
Best Local Similarity 34.8%; Pred. No. 2;  
Matches 23; Conservative 3; Mismatches 14; Indels 26; Gaps 4;  
Qy 36 RATATQAFPSI-TRAAPGARWGOQAHSPGLTAADGHEVFGIMFDAGSTGTRVHVQFTR 94  
Db 55 RALGTAAPFLDLRLDRPGLRW--DAHSG-GTAPDSHRI----- 89  
Qy 95 PPRETP 100  
Db 90 PPRES 95  
RESULT 7  
AAG00192  
XX AAG00192 standard; protein; 94 AA.  
XX  
AC AAG00192;  
XX  
XX 06-OCT-2000 (first entry)  
XX Human secreted protein, SEQ ID NO: 4273.  
XX  
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
XX gene therapy; chromosome mapping.  
XX  
XX Homo sapiens.  
XX

PN EPI033401-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 21-FEB-2000; 2000EP-00200610.  
XX  
PR 26-FEB-1999; 99US-0122487P.  
XX  
PA (GENSET ) GENSET.  
XX  
PI Dunas Milne Edwards J, Duclert A, Giordano J;  
XX  
DR WPI; 2000-500381/45.  
XX  
DR N-PSDB; AAC00198.  
XX  
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.  
XX  
PS Claim 13; SEQ ID NO 4273; 71bp + Sequence Listing; English.  
XX  
CC The present sequence is a polypeptide encoded by one of a large number of  
CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were  
CC prepared from total human RNAs or poly(A) RNAs derived from 30 different  
CC tissues. EST sequences usually correspond mainly to the 3' untranslated  
CC region (UTR) of the mRNA because they are often obtained from oligo-dT  
CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA  
CC sequences derived from the 5' ends of mRNAs and even in those cases where  
CC longer cDNA sequences have been obtained, the full 5' UTR is rarely  
CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can  
CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs  
CC are also used in diagnostic, forensic, gene therapy and chromosome  
CC mapping procedures. They are used to obtain upstream regulatory sequences  
CC and to design expression and secretion vectors  
XX  
SQ Sequence 94 AA;  
XX  
Query Match 9.8%; Score 63; DB 3; Length 94;  
Best Local Similarity 29.2%; Pred. No. 22;  
Matches 14; Conservative 12; Mismatches 20; Indels 2; Gaps 1;  
QY 72 VFYIMPDAGSTGRVHVFQTRPPREPTLTH--ETFKAVKPGLSAY 117  
Db 47 VKYGVIVDAGSSHTSLVYKVPKAKKENDTGVAHQVEECRVKPGISRKF 94  
XX  
RESULT 8  
AB078931  
ID ABB78931 standard; peptide; 92 AA.  
XX  
AC ABB78931;  
XX  
DT 30-JUL-2002 (first entry)  
XX  
DE Tumour necrosis factor-alpha binding amino acid sequence T14.08.  
XX  
KW Protein scaffold; antibody; binding protein; immunoglobulin;  
KW tumour necrosis factor alpha; TNF-alpha; protein framework.  
XX  
OS Homo sapiens.  
OS Synthetic.  
PN WO200232925-A2.  
XX  
XX 25-APR-2002.  
XX  
XX 16-OCT-2001; 2001WO-US032233.  
XX  
XX 16-OCT-2000; 2000US-00688566.  
XX  
XX (PHYL-) PHYLLOS INC.  
XX  
XX Lipovsek D, Wagner RW, Kuimelis RG;

XX  
DR WPI; 2002-444238/47.  
XX  
PT New non-antibody proteins having an immunoglobulin fold, useful in  
PT research, therapeutic or diagnostic fields, particularly as scaffolds for  
PT designing proteins with specific properties, e.g. for binding any antigen  
PT of interest.  
XX  
XX Claim 47; Fig 25; 94pp; English.  
XX  
CC The present invention describes a non-antibody protein, comprising a  
CC domain having an immunoglobulin-like fold, derived from a reference  
CC protein having a mutated amino acid sequence, where the non-antibody  
CC protein binds with a Kd at least as tight as 10 nM to a compound that is  
CC not bound as tightly by the reference protein. The non-antibody protein  
CC is useful as scaffolds for selecting or designing a protein framework  
CC with specific and favourable properties, e.g. for binding any antigen of  
CC interest, or for destroying or inactivating antibody molecules. The non-  
CC antibody protein is also useful in all areas where antibodies are used,  
CC e.g. research, therapeutic or diagnostic fields, and for screening novel  
CC binding proteins useful in the above-mentioned fields. The present  
CC proteins have thermodynamic properties superior to those of natural  
CC antibodies, and can be evolved rapidly in vitro. The present proteins or  
CC antibody mimics exhibit improved biophysical properties, such as  
CC stability under reducing conditions and solubility at high  
CC concentrations. In addition, these molecules are readily expressed and  
CC folded in prokaryotic systems (e.g. Escherichia coli), in eukaryotic  
CC systems (e.g. yeast), or in in vitro translation systems (e.g. rabbit  
CC reticulocyte lysate system). Furthermore, these proteins are extremely  
CC amenable to affinity maturation techniques involving multiple cycles of  
CC selection, e.g. in vitro selection using RNA-protein fusion technology,  
CC phage display or yeast display systems. The present sequence is used in  
CC the exemplification of the present invention  
XX  
SQ Sequence 92 AA;  
XX  
Query Match 9.7%; Score 62; DB 5; Length 92;  
Best Local Similarity 28.2%; Pred. No. 28;  
Matches 22; Conservative 8; Mismatches 24; Indels 24; Gaps 4;  
QY 37 ATATQAFPSITRAPDGAHQGQASHPLGTADGHEVFYIGIMPDAGSTGRVHVFQTRPP 96  
Db 13 ATPTSLXLS-----WXPRSH-----HRYRYRITY--GETGNSFVQGFVTP 52  
QY 97 RETPTLTHETFKAVKPG 114  
Db 53 WGTIA---TIDGLKPGV 66  
XX  
RESULT 9  
AAU66429  
ID AAU66429 standard; protein; 90 AA.  
XX  
AC AAU66429;  
XX  
DT 27-FEB-2002 (first entry)  
XX  
DE Propionibacterium acnes immunogenic protein #27325.  
XX  
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KW dermatological; osteopathic; neuroprotectant.  
XX  
XX Propionibacterium acnes.  
XX  
XX WO200181581-A2.  
XX  
XX 01-NOV-2001.  
XX  
XX 20-APR-2001; 2001WO-US012865.  
XX  
XX 21-APR-2000; 2000US-0199047P.

PR 02-JUN-2000; 2000US-0208841P.  
PR 07-JUL-2000; 2000US-0216747P.  
XX (CORI-) CORIXA CORP.  
XX  
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'Maisonneuve J, Zhang Y, Jen S, Carter D;  
XX WPI; 2001-616774/71.  
XX N-PSDB; AAS59733.  
XX  
XX Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris.  
XX  
XX Example 1; SEQ ID NO 27624; 1069pp; English.  
XX  
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 90 AA;  
SQ  
Query Match 9.6%; Score 61.5; DB 4; Length 90;  
Best Local Similarity 35.2%; Pred. No. 31;  
Matches 19; Conservative 3; Mismatches 27; Indels 5; Gaps 2;  
QY 36 RATATQAFSTTAAAPGARWQQQAHSPLGTAADGHEVFYIGIMFDAGTGTGRVHV 89  
Db 41 RMTSTLSASVDRRLP---WSAHPQPCRHAAQDHE--EPQCAQLKLGTRPHV 89  
RESULT 10  
ABM62948  
XX ABM62948 standard; protein; 90 AA.  
XX  
XX ABM62948;  
XX  
XX 20-OCT-2003 (first entry)  
XX  
XX Propionibacterium acnes predicted ORF-encoded polypeptide #27624.  
XX  
XX Acne vulgaris; antiseborrheic; dermatological; antibacterial;  
KW immunostimulant; immune response; vaccine.  
XX  
XX Propionibacterium acnes.  
OS  
XX WO2003033515-A1.  
PN  
XX 24-APR-2003.  
PD  
XX 11-OCT-2002; 2002WO-US032727.  
PF  
XX 15-OCT-2001; 2001US-00978825.  
PR  
XX (CORI-) CORIXA CORP.  
PA

XX Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;  
PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
PI Barth B, Vallieue-Douglas J;  
XX WPI; 2003-381789/36.  
DR N-PSDB; ACF64662.  
DR  
XX New Propionibacterium acnes polypeptides and polynucleotides encoding the  
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
PT or for stimulating an immune response specific for a P. acnes protein.  
XX  
XX Example 1; SEQ ID NO 27624; 1481pp; English.  
PS  
XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
CC encoding a Propionibacterium acnes protein. The invention also relates to  
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
CC immunogenic fragments of P. acnes polypeptides. The invention  
CC additionally encompasses expression vectors and host cells comprising a  
CC polynucleotide of the invention; antibodies against polypeptides of the  
CC invention; fusion proteins comprising a polypeptide of the invention; a  
CC method for stimulating an immune response specific for a P. acnes  
CC polypeptide and an isolated T cell population comprising T cells prepared  
CC via this method; a vaccine composition (comprising P. acnes polypeptides,  
CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
CC antigen-presenting cells that express the polypeptide); a method and kit  
CC for detecting or determining the presence or absence of P. acnes in a  
CC patient; and a method for inhibiting the development of P. acnes in a  
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
CC proteins, T cell populations or antigen-presenting cells that express the  
CC polypeptides are useful for diagnosing, preventing or treating acne  
CC vulgaris, or for stimulating an immune response specific for a P. acnes  
CC protein. The polynucleotides can also be used as probes or primers for  
CC nucleic acid hybridisation. The vaccine composition is useful for the  
CC stimulation of an immune response against P. acnes, or for treating acne,  
CC and the kit is useful for performing a diagnostic assay. The present  
CC sequence represents a polypeptide predicted to be encoded by an ORF (open  
CC reading frame) contained within the P. acnes polynucleotides of the  
CC invention. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 90 AA;  
SQ  
Query Match 9.6%; Score 61.5; DB 6; Length 90;  
Best Local Similarity 35.2%; Pred. No. 31;  
Matches 19; Conservative 3; Mismatches 27; Indels 5; Gaps 2;  
QY 36 RATATQAFSTTAAAPGARWQQQAHSPLGTAADGHEVFYIGIMFDAGTGTGRVHV 89  
Db 41 RMTSTLSASVDRRLP---WSAHPQPCRHAAQDHE--EPQCAQLKLGTRPHV 89  
RESULT 11  
ADD26584  
ID ADD26584 standard; protein; 100 AA.  
XX  
XX ADD26584;  
XX  
XX 15-JAN-2004 (first entry)  
DT  
XX BGS-19 immunoglobulin-like domain 1 SEQ ID NO:9.  
DE  
XX human; cell surface protein; immunoglobulin; BGS-19; cytostatic;  
KW gynaecological; immunosuppressive; antiinflammatory; antiasthmatic;  
KW antidiabetic; dermatological; gene therapy.  
XX  
XX Homo sapiens.  
OS  
XX WO2003083078-A2.  
PN  
XX 09-OCT-2003.  
PD  
XX (CORI-) CORIXA CORP.  
PA

PF 28-MAR-2003; 2003WO-US009676.  
 XX  
 PR 28-MAR-2002; 2002US-0368422P.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Lee LM, Feder JN, Siemers NO, Wu S, Chen J;  
 XX  
 DR WPI; 2003-804052/75.  
 XX  
 PT New isolated nucleic acid molecule encoding BGS-19 polypeptides, useful  
 PT for preventing, treating or ameliorating a medical condition, such as a  
 PT disorder related to aberrant immunoglobulin cell surface receptor  
 PT activity.  
 XX  
 PS Disclosure; SEQ ID NO 9; 224pp; English.  
 XX  
 CC The present invention describes human cell surface protein with  
 CC immunoglobulin folds, designated BGS-19 (I). (I) has cytostatic,  
 CC synaescological, immunosuppressive, antiinflammatory, antiasthmatic,  
 CC antidiabetic and dermatological activities, and can be used in gene  
 CC therapy. (I) can be used for preventing, treating or ameliorating a  
 CC medical condition, such as a disorder related to aberrant immunoglobulin  
 CC cell surface receptor activity; a cellular adhesion disorder; a disorder  
 CC related to hyper- or hypo-immunoglobulin receptor activity; a disorder  
 CC related to aberrant signal transduction; a reproductive disorder; a  
 CC female reproductive disorder; an ovarian disorder; ovarian cancer;  
 CC dysfunctional uterine bleeding; amenorrhoea; primary dysmenorrhoea;  
 CC sexual dysfunction; infertility; pelvic inflammatory disease;  
 CC endometriosis; placental aromatase deficiency; premature menopause;  
 CC placental dysfunction; hormone deficiency; oestrogen deficiency; aberrant  
 CC androgen metabolism; aberrant onset of female puberty; aberrant showing  
 CC of female primary sexual characteristics; aberrant showing of female  
 CC secondary sexual characteristics; precocious puberty; precocious  
 CC pseudopuberty; incomplete isosexual precocity; premature thelarche;  
 CC premature adrenarche; premature pubarche; polycystic ovarian disease;  
 CC aberrant ovarian cycle; menorrhagia; metrorrhagia; menometrorrhagia;  
 CC dysmenorrhoea; hypomenorrhoea; polymenorrhoea; dysfunctional uterine  
 CC bleeding; resistant-ovary syndrome; hermaphroditism; immune disorders;  
 CC inflammatory disorders; arthritis; asthma; immunodeficiency diseases such  
 CC as AIDS; leukaemia; rheumatoid arthritis; granulomatous disease;  
 CC inflammatory bowel disease; sepsis; acne; neutropenia; neutrophilia;  
 CC psoriasis; hypersensitivity; such as T-cell mediated cytotoxicity;  
 CC immune reactions to transplanted organs and tissues; such as host-versus-  
 CC graft and graft-versus-host diseases; or autoimmunity disorders;  
 CC autoimmune infertility; Addison's Disease; haemolytic anaemia;  
 CC antiphospholipid syndrome; rheumatoid arthritis; dermatitis; allergic  
 CC encephalomyelitis; glomerulonephritis; Goodpasture's Syndrome; Graves'  
 CC Disease; Multiple Sclerosis; Myasthenia Gravis; Neuritis; Ophthalmia;  
 CC Bullous Pemphigoid; Pemphigus; Polyendocrinopathies; Purpura; Reiter's  
 CC Disease; Stiff-Man Syndrome; autoimmune thyroiditis; Systemic Lupus  
 CC Erythematosus; Autoimmune Pulmonary Inflammation; Guillain-Barre Syndrome  
 CC ; Insulin dependent diabetes mellitus; autoimmune inflammatory eye  
 CC disease; lens tissue injury; demyelination; systemic lupus erythematosus;  
 CC drug induced haemolytic anaemia; rheumatoid arthritis; Sjogren's disease;  
 CC and scleroderma. The present sequence is used in the exemplification of  
 CC the present invention.  
 XX  
 SQ Sequence 100 AA;  
 Query Match 9.5%; Score 60.5; DB 7; Length 100;  
 Best Local Similarity 29.7%; Pred. No. 47;  
 Matches 27; Conservative 12; Mismatches 33; Indels 19; Gaps 5;  
 QY 9 SLRYAK-VAAPDLGCVGFIVVAYIK--WRATATQAFSITRAAPARWGQOAHSLIGT 65  
 DB 10 SLQGRQVPYPRFEGICVVSCHLSYPRDQWDESTLAIVYWKRTSP-----KTGAPVAT 63  
 QY 66 AADGHEVFYGI MPDAGSTGRVHVFOFTRP 96  
 DB 64 NNQGRV-----EMSTRDR---FOLTGDP 84

RESULT 12  
 ID AU31171  
 XX AU31171 standard; protein; 114 AA.  
 XX  
 AC AU31171;  
 XX  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Novel human secreted protein #1662.  
 XX  
 KM Human; vaccination; gene therapy; nutritional supplement;  
 KM stem cell proliferation; haematopoiesis; nerve tissue regeneration;  
 KM immune suppression; immune stimulation; anti-inflammatory; leukaemia.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200179449-A2.  
 XX  
 PD 25-OCT-2001.  
 XX  
 PF 16-APR-2001; 2001WO-US008656.  
 XX  
 PR 18-APR-2000; 2000US-00552929.  
 PR 26-JAN-2001; 2001US-00770160.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Dermanac RT;  
 XX  
 DR WPI; 2001-611725/70.  
 XX  
 PT Nucleic acids encoding a range of human polypeptides, useful in genetic  
 PT vaccination, testing and therapy.  
 XX  
 PS Claim 20; Page 413; 765pp; English.  
 XX  
 CC The invention relates to novel human secreted polypeptides. The  
 CC polypeptides and antibodies to the polypeptides are useful for  
 CC determining the presence of or predisposition to a disease associated  
 CC with altered levels of polypeptide. The polypeptides are also useful for  
 CC identifying agents (agonists and antagonists) that bind to them. Cells  
 CC expressing the proteins are useful for identifying a therapeutic agent  
 CC for use in treatment of a pathology related to aberrant expression or  
 CC physiological interactions of the polypeptide. Vectors comprising the  
 CC nucleic acids encoding the polypeptides and cells genetically engineered  
 CC to express them are also useful for producing the proteins. The proteins  
 CC are useful in genetic vaccination, testing and therapy, and can be used  
 CC as nutritional supplements. They may be used to increase stem cell  
 CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon  
 CC and/or nerve tissue growth or regeneration; immune suppression and/or  
 CC stimulation; as anti-inflammatory agents; and in treatment of leukemias.  
 CC AAU93510-AU93304 represent the amino acid sequences of novel human  
 CC secreted proteins of the invention  
 XX  
 SQ Sequence 114 AA;  
 Query Match 9.1%; Score 58.5; DB 4; Length 114;  
 Best Local Similarity 27.0%; Pred. No. 95;  
 Matches 30; Conservative 10; Mismatches 22; Indels 49; Gaps 7;  
 QY 30 AYIKWRATATQAFSITRAAPARWGQOAHSLIGTAAAGHEVFYGIWFDAGSTGRVH- 88  
 DB 19 AGVCHMDRLRSIQ-----PPGSR-----APASRV-----AGTTGTRHHA 56  
 QY 89 -VPOF-----TRPREPTLTHETFAVYKGLS 115  
 DB 57 RLIFFFPPFFFLVETGFHAVSEDDDLFTSGDLPIITLSQS--AGIPGMS 105  
 RESULT 13  
 ID ABB40819  
 XX ABB40819 standard; peptide; 82 AA.



```
AC ABB40819;
XX 04-FEB-2002 (first entry)
XX DE Peptide #8325 encoded by human foetal liver single exon probe.
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX OS Homo sapiens.
XX PN WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000669.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WIPI; 2001-483447/52.
XX DR Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human fetal liver.
XX PS Claim 27; SEQ ID NO 33454; 639pp + Sequence Listing; English.
XX CC The invention relates to a single exon nucleic acid probe for measuring
XX CC single exon gene expression in a sample derived from human foetal liver. The
XX CC single exon nucleic acid probes may be used for predicting, measuring and
XX CC displaying gene expression in samples derived from human fetal liver. The
XX CC present sequence is a peptide encoded by a single exon nucleic acid probe
XX CC of the invention. Note: The sequence data for this patent did not form
XX CC part of the printed specification, but was obtained in electronic format
XX CC directly from WIPO at ftp.wipo.int/pub/published_pat_sequences
XX SQ Sequence 82 AA;
XX
XX Query Match 9.1%; Score 58; DB 4; Length 82;
XX Best Local Similarity 30.9%; Pred. No. 71;
XX Matches 21; Conservative 9; Mismatches 18; Indels 20; Gaps 4;
XX
XX QY 40 TQAFPSI-TRAAPCARWG-QQAHSPLGTAADGHEVFGIMFDAGSTGTRVHVQFTRPPR 97
XX DB 17 TQGLGIGQKSAPRAPWALKVAQGPLEPQA-----PKVN-RETTTRPPR 58
XX
XX QY 98 ETPTLTHE 105
XX DB 59 KSPSLPQE 66
XX
XX Query Match 9.1%; Score 58; DB 4; Length 82;
XX Best Local Similarity 30.9%; Pred. No. 71;
XX Matches 21; Conservative 9; Mismatches 18; Indels 20; Gaps 4;
XX
XX QY 40 TQAFPSI-TRAAPCARWG-QQAHSPLGTAADGHEVFGIMFDAGSTGTRVHVQFTRPPR 97
XX DB 17 TQGLGIGQKSAPRAPWALKVAQGPLEPQA-----PKVN-RETTTRPPR 58
XX
XX QY 98 ETPTLTHE 105
XX DB 59 KSPSLPQE 66
XX
XX RESULT 14
XX AAM34585
XX ID AAM34585 standard; protein; 82 AA.
XX AC AAM34585;
XX
XX DT 17-OCT-2001 (first entry)
XX DE Peptide #8622 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder.
XX OS Homo sapiens.
```

```
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000663.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WIPI; 2001-488897/53.
XX DR Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human placenta.
XX PS Claim 27; SEQ ID NO 34854; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENP:
XX CC see AAI31315-AAJ57546). The present sequence is a peptide encoded by one
XX CC such probe. The probes are useful for producing a microarray for
XX CC predicting, measuring and displaying gene expression in samples derived
XX CC from human placenta. The probes are useful for antenatal diagnosis of
XX CC human genetic disorders
XX SQ Sequence 82 AA;
XX
XX Query Match 9.1%; Score 58; DB 4; Length 82;
XX Best Local Similarity 30.9%; Pred. No. 71;
XX Matches 21; Conservative 9; Mismatches 18; Indels 20; Gaps 4;
XX
XX QY 40 TQAFPSI-TRAAPCARWG-QQAHSPLGTAADGHEVFGIMFDAGSTGTRVHVQFTRPPR 97
XX DB 17 TQGLGIGQKSAPRAPWALKVAQGPLEPQA-----PKVN-RETTTRPPR 58
XX
XX QY 98 ETPTLTHE 105
XX DB 59 KSPSLPQE 66
XX
XX RESULT 15
XX AAM74471
XX ID AAM74471 standard; protein; 82 AA.
XX AC AAM74471;
XX
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 34777.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX OS Homo sapiens.
XX PN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000668.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
```

PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
VV

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488900/53.

PT Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human bone marrow.  
VV

PS Example 4; SEQ ID NO 34777; 658pp + Sequence Listing; English.  
xy

CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
CC protein encoded by one of the probes of the invention

Sequence 82 AA;

Query Match	9.1%	Score 58;	DB 4;	Length 82;
Best Local Similarity	30.9%	Pred. No. 71;		
Matches 21; Conservative	9;	Mismatches 18;	Indels 20;	Gaps 4;

QY 40 TQAFESI-TRAPGARNG-QQAHSPLETAADGHEVFYGLMEADASTGTRVHVFOGTRPPR 97  
|||::|||:|||||:|:|||||  
Db 17 TQGFGLIQEKAPRAPAPALKVAQGLEPQA-----PKYN-RETTWRPPR 58

QY 98 ETPTLTHE 105

Db 59 KSPSLPQE 66

Search completed: June 8, 2004, 10:00:18  
Job time : 57 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 10:01:46 ; Search time 42 Seconds  
(without alignments)  
803,825 Million cell updates/sec

Title: US-09-905-589a-2\_COPY\_1\_120

Perfect score: 640  
Sequence: 1 MRKSNHGSRLKAVKAYPLG.....TLTHETPKAYKPGLSAYADD 120

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 526131

Minimum DB seq length: 0  
Maximum DB seq length: 120

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PC1\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PC1S\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10C\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78	12.2	14	US-10-286-926-44	Sequence 44, App1
2	78	12.2	108	US-10-424-599-217882	Sequence 217882,
3	66.5	10.4	119	US-10-424-599-219167	Sequence 219167,
4	66	10.3	80	US-10-062-831-220	Sequence 220, App
5	65	10.3	80	US-10-062-831-220	Sequence 220, App
6	63	9.8	61	US-10-424-599-282823	Sequence 282823,
7	60.5	9.5	100	US-10-403-938-9	Sequence 9, App1
8	59	9.2	81	US-10-424-599-230474	Sequence 230474,
9	58	9.1	82	US-09-864-761-44698	Sequence 44698, A
10	58	9.1	91	US-10-424-599-154715	Sequence 154715,
11	57.5	9.0	112	US-10-264-237-1834	Sequence 1834, Ap
12	57.5	9.0	118	US-10-108-260A-2613	Sequence 2613, Ap
13	57	8.9	63	US-10-424-599-165371	Sequence 165371,
14	57	8.9	72	US-09-864-761-33368	Sequence 33368, A
15	57	8.9	108	US-09-738-626-5443	Sequence 5443, Ap

#### ALIGNMENTS

16	56.5	8.8	81	12	US-10-424-599-255385	Sequence 255385,
17	56.5	8.8	87	12	US-10-424-599-172253	Sequence 172253,
18	56.5	8.8	119	12	US-10-425-114-57544	Sequence 57544, A
19	56	8.8	52	12	US-10-424-599-181841	Sequence 181841,
20	55.5	8.7	58	9	US-09-835-147-31	Sequence 31, App1
21	55	8.6	101	15	US-10-108-260A-4537	Sequence 4537, Ap
22	54.5	8.5	61	12	US-10-424-599-160185	Sequence 160185,
23	54.5	8.5	89	12	US-10-424-599-218455	Sequence 218455,
24	54	8.4	92	12	US-10-424-599-218182	Sequence 218182,
25	53.5	8.4	70	9	US-09-867-550-364	Sequence 364, App
26	53.5	8.4	77	15	US-10-264-049-2513	Sequence 2513, Ap
27	53	8.3	82	12	US-10-424-599-201768	Sequence 201768,
28	53	8.3	107	15	US-10-104-047-3230	Sequence 3230, Ap
29	53	8.3	112	12	US-10-424-599-282912	Sequence 282912,
30	52.5	8.2	54	14	US-10-083-357-1261	Sequence 1261, Ap
31	52.5	8.2	64	12	US-10-424-599-250977	Sequence 250977,
32	52.5	8.2	86	12	US-10-001-885-166	Sequence 166, App
33	52.5	8.2	96	14	US-10-302-456-1	Sequence 1,
34	52.5	8.2	116	12	US-10-282-122A-62053	Sequence 62053, A
35	52.5	8.2	117	12	US-10-424-599-192570	Sequence 192570,
36	52	8.1	78	12	US-10-424-599-236196	Sequence 236196,
37	52	8.1	98	12	US-10-424-599-239102	Sequence 239102,
38	52	8.1	103	14	US-10-424-599-279416	Sequence 279416,
39	52	8.1	110	14	US-10-029-386-29859	Sequence 29859, A
40	51.5	8.0	73	12	US-10-424-599-274655	Sequence 274655,
41	51.5	8.0	82	9	US-09-864-761-39751	Sequence 39751, A
42	51	8.0	73	12	US-10-424-599-229454	Sequence 229454,
43	51	8.0	87	12	US-10-424-599-171475	Sequence 171475,
44	51	8.0	111	12	US-10-424-599-277082	Sequence 277082,
45	51	8.0	113	12	US-10-424-599-274427	Sequence 274427,

RESULT 1  
US-10-286-926-44  
Sequence 44, Application US/10286926  
Publication No. US20030175752A1  
GENERAL INFORMATION:  
APPLICANT: Ford, John  
APPLICANT: Malec, John  
TITLE OF INVENTION: Methods and Materials Relating to CD39-Like  
FILE REFERENCE: 28110/36457CON  
CURRENT APPLICATION NUMBER: US/10/266, 926  
CURRENT FILING DATE: 2002-11-01  
PRIOR APPLICATION NUMBER: 09/557, 800  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/481, 238  
PRIOR FILING DATE: 2000-01-11  
PRIOR APPLICATION NUMBER: 09/370, 265  
PRIOR FILING DATE: 1999-08-09  
PRIOR APPLICATION NUMBER: PCT/US99/16180  
PRIOR FILING DATE: 1999-07-16  
PRIOR APPLICATION NUMBER: 09/350836  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: 09/273447  
PRIOR FILING DATE: 1999-03-19  
PRIOR APPLICATION NUMBER: 09/122449  
PRIOR FILING DATE: 1998-07-24  
PRIOR APPLICATION NUMBER: 09/244444  
PRIOR FILING DATE: 1999-02-04  
PRIOR APPLICATION NUMBER: 09/118, 205  
PRIOR FILING DATE: 1998-07-16  
NUMBER OF SEQ ID NOS: 54  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 44  
LENGTH: 14  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: peptide  
US-10-286-926-44

Query Match 12.2%; Score 78; DB 14; Length 14;  
Best Local Similarity 100.0%; Pred. No. 0.012;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 TRPPRETPLTHET 106  
| | | | | | | | | | | | | | | |  
Db 1 TRPPRETPLTHET 14

## RESULT 2

US-10-424-599-217882  
; Sequence 217882, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 217882  
; LENGTH: 108  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_38774C.1.pap  
US-10-424-599-217882

Query Match 12.2%; Score 78; DB 12; Length 108;  
Best Local Similarity 68.4%; Pred. No. 0.15;  
Matches 13; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 74 YGIMFDAGSTGRVHFQF 92  
| | | | | | | | | | | | | | | |  
Db 68 YAVIFDAGSGSRVHFHF 86

## RESULT 3

US-10-424-599-219167  
; Sequence 219167, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 219167  
; LENGTH: 119  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)...(119)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_39935C.1.pap  
US-10-424-599-219167

Query Match 10.4%; Score 66.5; DB 12; Length 119;  
Best Local Similarity 37.0%; Pred. No. 4.6;

Matches 17; Conservative 5; Mismatches 23; Indels 1; Gaps 1;  
QY 76 IMFADAGSTGRVHFQFTRPPRETPT-LTHETPKAVKGLSAYADD 120  
| | | | | | | | | | | | | | | |  
Db 28 VIFDGGTGRVHFHQAQNDLLPMGVGLNKRITPGFXAYGVD 73

## RESULT 4

US-10-062-831-220  
; Sequence 220, Application US/10062831  
; Publication No. US20030105297A1  
; GENERAL INFORMATION:  
; APPLICANT: Steven M. Ruben, et al.  
; TITLE OF INVENTION: 32 Human Secreted Proteins  
; FILE REFERENCE: P2006P1  
; CURRENT APPLICATION NUMBER: US/10/062,831  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: 09/690,454  
; PRIOR FILING DATE: 1998-11-10  
; PRIOR APPLICATION NUMBER: PCT/US98/10868  
; PRIOR FILING DATE: May 28, 1998  
; PRIOR APPLICATION NUMBER: 60/044,039  
; PRIOR FILING DATE: May 30, 1997  
; PRIOR APPLICATION NUMBER: 60/048,093  
; PRIOR FILING DATE: May 30, 1997  
; PRIOR APPLICATION NUMBER: 60/048,190  
; PRIOR FILING DATE: May 30, 1997  
; PRIOR APPLICATION NUMBER: 60/050,935  
; PRIOR FILING DATE: May 30, 1997  
; PRIOR APPLICATION NUMBER: 60/048,101  
; PRIOR FILING DATE: May 30, 1997  
; PRIOR APPLICATION NUMBER: 60/048,356  
; PRIOR FILING DATE: May 30, 1997  
; PRIOR APPLICATION NUMBER: 60/056,250  
; PRIOR FILING DATE: August 29, 1997  
; PRIOR APPLICATION NUMBER: 60/056,296  
; PRIOR FILING DATE: August 29, 1997  
; PRIOR APPLICATION NUMBER: 60/056,293  
; PRIOR FILING DATE: August 29, 1997  
; NUMBER OF SEQ ID NOS: 229  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 220  
; LENGTH: 80  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (72)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-10-062-831-220

Query Match 10.3%; Score 66; DB 14; Length 80;  
Best Local Similarity 57.9%; Pred. No. 3.2;  
Matches 11; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 51 PGARWGQQAHSPLGTAADG 69  
| | | | | | | | | | | | | | | |  
Db 40 PRIRWGQEAHVPAQAQEG 58

## RESULT 5

US-10-062-599-220  
; Sequence 220, Application US/10062599  
; Publication No. US20030195346A1  
; GENERAL INFORMATION:  
; APPLICANT: Steven M. Ruben, et al.  
; TITLE OF INVENTION: 32 Human Secreted Proteins  
; FILE REFERENCE: P2006P1  
; CURRENT APPLICATION NUMBER: US/10/062,599  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: 09/690,454  
; PRIOR FILING DATE: 2000-10-18  
; PRIOR APPLICATION NUMBER: 09/189,144





Tue Jun 8 12:10:42 2004

us-09-905-589a-2\_copy\_1\_120.rapp

Page 5

```

RESULT 12
US-10-108-260A-2613
; Sequence 2613, Application US//10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELEX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2613
; LENGTH: 118
; TYPE: PR1
; ORGANISM: Homo sapiens
US-10-108-260A-2613

```

```

RESULT 15
US-09-738-626-5443
; Sequence 5443, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIALI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5443
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5443

Query Match      8.9%; Score 57; DB 9; Length 108;
Best Local Similarity 28.8%; Pred. No. 61;
Matches 19; Conservative 5; Mismatches 42; Indels 0; Gaps 0;

Qy 53 ARWQQAHSPIGTAADGHEVEYGYIMFDAGSTGTRVHVFOFTRPPRETPTLTHTETPKAVKP 112
Db 20 SRWNKFAVSPIGFAGALRFLGFSAAASTPSSVALVLMVASPSRSASALIASSSKASVS 79

Qy 113 GLSAYA 118
Db 80 GLYSVA 85

```

Search completed: June 8, 2004, 10:07:25  
Job time : 43 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 09:59:16 ; Search time 22 Seconds  
(without alignments)  
281.596 Million cell updates/sec

Title: US-09-905-589A-2\_COPY\_1\_120  
Perfect score: 640  
Sequence: 1 MKRISNHSGLRVAKVAYPLG.....TLTHTFFAKVQGLSAVAD 120

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 277640

Minimum DB seq length: 0  
Maximum DB seq length: 120

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/6B\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78	12.2	14	US-09-608-285A-44	Sequence 44, Appl
2	78	12.2	14	US-09-557-800C-44	Sequence 44, Appl
3	67.5	10.5	96	US-09-252-991A-27819	Sequence 27819, A
4	66	10.3	80	US-09-690-454-220	Sequence 220, App
5	54	8.4	108	US-08-466-033-106	Sequence 106, App
6	54	8.4	108	US-08-444-733-106	Sequence 106, App
7	54	8.4	108	US-08-464-134-106	Sequence 106, App
8	54	8.4	108	US-08-461-161-106	Sequence 106, App
9	54	8.4	108	US-08-485-910-106	Sequence 106, App
10	54	8.4	108	PCT-US95-06266-89	Sequence 89, Appl
11	53	8.0	109	US-09-543-681A-6609	Sequence 6609, Ap
12	51.5	8.0	113	US-09-489-039A-8242	Sequence 8242, Ap
13	51	8.0	101	US-09-489-039A-12049	Sequence 12049, A
14	51	8.0	108	US-08-959-812-10	Sequence 10, Appl
15	50	7.8	90	US-09-621-976-4538	Sequence 4538, Ap
16	50	7.8	90	US-09-621-976-4538	Sequence 206, Appl
17	49	7.7	67	US-09-230-041-26	Sequence 26, Appl
18	49	7.7	70	US-09-079-372-2	Sequence 2, Appl
19	49	7.7	100	US-09-079-372-14	Sequence 14, Appl
20	48.5	7.5	116	US-09-489-039A-11188	Sequence 11188, A
21	48	7.5	52	US-09-621-976-5873	Sequence 5873, Ap
22	48	7.5	96	US-09-369-247-72	Sequence 72, Appl
23	48	7.5	112	US-09-621-976-5677	Sequence 5677, Ap
24	48	7.5	112	US-09-621-976-5805	Sequence 5805, Ap
25	47.5	7.4	40	US-08-469-260A-467	Sequence 467, App
26	47.5	7.4	40	US-08-469-260A-467	Sequence 467, App
27	47.5	7.4	40	US-08-467-344A-467	Sequence 467, App

28	47.5	7.4	83	US-09-252-991A-19508	Sequence 19508, A
29	47.5	7.4	87	US-09-205-258-894	Sequence 894, App
30	47.5	7.4	114	US-08-165-754-1	Sequence 1, Appl
31	47.5	7.4	114	US-07-845-5925-1	Sequence 1, Appl
32	47	7.3	62	US-09-621-976-7044	Sequence 7044, Appl
33	47	7.3	68	US-09-107-532A-3872	Sequence 3872, Ap
34	47	7.3	72	US-09-540-336-2170	Sequence 2170, Ap
35	47	7.3	89	US-09-489-039A-13860	Sequence 13860, A
36	47	7.3	114	US-09-252-991A-25775	Sequence 25775, A
37	47	7.3	114	US-09-489-039A-9698	Sequence 9698, Ap
38	47	7.3	117	US-09-252-991A-31460	Sequence 31460, A
39	46.5	7.3	73	US-08-252-991A-20704	Sequence 20704, A
40	46.5	7.3	93	US-09-328-352-5221	Sequence 3221, Ap
41	46.5	7.3	94	US-09-621-976-5714	Sequence 5714, Ap
42	46.5	7.3	107	US-09-621-976-5595	Sequence 5595, Ap
43	46.5	7.3	117	US-09-134-000C-3881	Sequence 3881, Ap
44	46	7.2	25	US-07-752-101A-6	Sequence 6, Appl
45	46	7.2	33	US-07-741-453A-12	Sequence 12, Appl

#### ALIGNMENTS

```
RESULT 1
US-09-608-285A-44
; Sequence 44, Application US/09608285A
; Patent No. 6335013
; GENERAL INFORMATION:
; APPLICANT: Mulero, John
; APPLICANT: Ford, Julio
; APPLICANT: Young, George
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO CD39-LIKE
; FILE REFERENCE: 28110/36570
; CURRENT APPLICATION NUMBER: US/09/608, 285A
; PRIOR APPLICATION NUMBER: 09/583, 231
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 09/557, 800
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/481, 238
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: 09/370, 265
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: PCT/US99/16180
; PRIOR FILING DATE: 1999-07-16
; PRIOR APPLICATION NUMBER: 09/350, 836
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/273, 447
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 09/244, 444
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: 09/122, 449
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 09/118, 205
; PRIOR FILING DATE: 1998-07-16
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 44
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: peptide
US-09-608-285A-44

Query Match      12.2%  Score 78;  DB 4;  Length 14;
Best Local Similarity 100.0%;  Pred. No. 0.0018;
Matches 14;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY      93 TRPREPTLTHTT 106
DB      1 TRPREPTLTHTT 14
```

Best Local Similarity 29.1%; Pred. No. 0.54; Mismatches 13; Indels 17; Gaps 4;  
Matches 23; Conservative 13; Mismatches 26; Indels 17; Gaps 4;  
QY 49 AAGCARGQQAHSPGLGTAADGHEVFGIMFDAGSTGTRVHVFGFTPRPPETPLTHE--- 105  
Db 21 ALPGR--GRISSAPLGATSGSLSHGL-----TRRYLTGNGSEENPTLPDQSPG 70  
QY 106 --TPKAVK--PGLSAYADD 120  
Db 71 DRTYRERRRPFGCPAGTDD 89  
RESULT 4  
US-09-690-454-220  
; Sequence 220, Application US/09690454  
; Patent No. 6531447  
; GENERAL INFORMATION:  
; APPLICANT: Steven M. Ruben, et al.  
; TITLE OF INVENTION: 32 Human Secreted Proteins  
; FILE REFERENCE: P2006P1  
; CURRENT APPLICATION NUMBER: US/09/690,454  
; CURRENT FILING DATE: 2000-10-18  
; PRIOR APPLICATION NUMBER: 09/189,144  
; PRIOR FILING DATE: 1998-11-10  
; PRIOR APPLICATION NUMBER: 60/044,039  
; PRIOR FILING DATE: May 30, 1997  
; PRIOR APPLICATION NUMBER: 60/048,093  
; PRIOR FILING DATE: May 30, 1997  
; PRIOR APPLICATION NUMBER: 60/048,190  
; PRIOR FILING DATE: May 30, 1997  
; PRIOR APPLICATION NUMBER: 60/050,935  
; PRIOR FILING DATE: May 30, 1997  
; PRIOR APPLICATION NUMBER: 60/048,101  
; PRIOR FILING DATE: May 30, 1997  
; PRIOR APPLICATION NUMBER: 60/048,356  
; PRIOR FILING DATE: May 30, 1997  
; PRIOR APPLICATION NUMBER: 60/056,250  
; PRIOR FILING DATE: August 29, 1997  
; PRIOR APPLICATION NUMBER: 60/056,296  
; PRIOR FILING DATE: August 29, 1997  
; PRIOR APPLICATION NUMBER: 60/056,293  
; PRIOR FILING DATE: August 29, 1997  
; NUMBER OF SEQ ID NOS: 229  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 220  
; LENGTH: 80  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (72)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-690-454-220  
Query Match 10.3%; Score 66; DB 4; Length 80;  
Best Local Similarity 57.9%; Pred. No. 0.66; Mismatches 2; Indels 6; Gaps 0;  
Matches 11; Conservative 2; Mismatches 6; Indels 6; Gaps 0;  
QY 51 PGARWGQQAHSPGLGTAADG 69  
Db 40 PRIRWGQEAHVPAQAQEG 58  
RESULT 5  
US-08-466-033-106  
; Sequence 106, Application US/08466033  
; Patent No. 5766840  
; GENERAL INFORMATION:  
; APPLICANT: Kim, Jungshuh P.  
; APPLICANT: Wages, John  
; APPLICANT: Young, LaVonne M.  
; APPLICANT: Fry, Kirk E.  
; APPLICANT: Linnen, Jeffrey M.

RESULT 2  
US-09-557-800C-44  
; Sequence 44, Application US/09557800C  
; Patent No. 6476211  
; GENERAL INFORMATION:  
; APPLICANT: Ford, John  
; APPLICANT: Mulero, Julio  
; APPLICANT: Yeung, George  
; TITLE OF INVENTION: Methods and Materials Relating to CD39-Like  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 28110/36457  
; CURRENT APPLICATION NUMBER: US/09/557,800C  
; CURRENT FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/481,238  
; PRIOR FILING DATE: 2000-01-11  
; PRIOR APPLICATION NUMBER: 09/370,265  
; PRIOR FILING DATE: 1999-08-09  
; PRIOR APPLICATION NUMBER: PCT/US99/16180  
; PRIOR FILING DATE: 1999-07-16  
; PRIOR APPLICATION NUMBER: 09/350836  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: 09/273447  
; PRIOR FILING DATE: 1999-03-19  
; PRIOR APPLICATION NUMBER: 09/122449  
; PRIOR FILING DATE: 1998-07-24  
; PRIOR APPLICATION NUMBER: 09/244444  
; PRIOR FILING DATE: 1999-02-04  
; PRIOR APPLICATION NUMBER: 09/118,205  
; PRIOR FILING DATE: 1998-07-16  
; NUMBER OF SEQ ID NOS: 56  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 44  
; LENGTH: 14  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: peptide  
US-09-557-800C-44

Query Match 12.2%; Score 78; DB 4; Length 14;  
Best Local Similarity 100.0%; Pred. No. 0.0018; Mismatches 0; Indels 0; Gaps 0;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 93 TRPPRETPLTHET 106  
Db 1 TRPPRETPLTHET 14

RESULT 3  
US-09-252-991A-27819  
; Sequence 27819, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 27819  
; LENGTH: 96  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-27819

Query Match 10.5%; Score 67.5; DB 4; Length 96;

TITLE OF INVENTION: Hepatitis G Virus and Molecular  
TITLE OF INVENTION: Cloning Thereof  
NUMBER OF SEQUENCES: 277  
CORRESPONDENCE ADDRESS:  
ADDRESS: Dellinger & Associates  
STREET: 350 Cambridge Ave., Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/466,033  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/389,886  
FILING DATE: 15-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/357,509  
FILING DATE: 16-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,729  
FILING DATE: 26-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/344,271  
FILING DATE: 23-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/285,558  
FILING DATE: 03-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/285,543  
FILING DATE: 03-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/246,985  
FILING DATE: 20-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Fadian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 106:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 108 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-466-033-106

Query Match 8.4%; Score 54; DB 1; Length 108;  
Best Local Similarity 38.2%; Pred. No. 33;  
Matches 13; Conservative 6; Mismatches 13; Indels 2; Gaps 1;

RESULT 6  
US-08-444-733-106  
Sequence 106, Application US/08444733  
Patent No. 5824507  
GENERAL INFORMATION:  
APPLICANT: Kim, Jungshuh P.  
APPLICANT: Wages, John  
APPLICANT: Young, Lavonne M.  
APPLICANT: Fry, Kirk E.

APPLICANT: Limen, Jeffrey M.  
TITLE OF INVENTION: Hepatitis G Virus and Molecular  
TITLE OF INVENTION: Cloning Thereof  
NUMBER OF SEQUENCES: 277  
CORRESPONDENCE ADDRESS:  
ADDRESS: Dellinger & Associates  
STREET: 350 Cambridge Ave., Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/444,733  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/389,886  
FILING DATE: 15-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/357,509  
FILING DATE: 16-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,729  
FILING DATE: 26-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/344,271  
FILING DATE: 23-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/285,558  
FILING DATE: 03-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/285,543  
FILING DATE: 03-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/246,985  
FILING DATE: 20-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Fadian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 106:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 108 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-444-733-106

Query Match 8.4%; Score 54; DB 2; Length 108;  
Best Local Similarity 38.2%; Pred. No. 33;  
Matches 13; Conservative 6; Mismatches 13; Indels 2; Gaps 1;

RESULT 7  
US-08-464-134-106  
Sequence 106, Application US/08464134  
Patent No. 5849532  
GENERAL INFORMATION:  
APPLICANT: Kim, Jungshuh P.  
APPLICANT: Wages, John  
APPLICANT: Young, Lavonne M.

```

; APPLICANT: Fry, Kirk E.
; APPLICANT: Limen, Jeffrey M.
; TITLE OF INVENTION: Hepatitis G Virus and Molecular
; Cloning Thereof
; NUMBER OF SEQUENCES: 277
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave., Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,134
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/389,886
; FILING DATE: 15-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/357,509
; FILING DATE: 16-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,729
; FILING DATE: 26-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/344,271
; FILING DATE: 23-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/285,558
; FILING DATE: 03-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/285,543
; FILING DATE: 03-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/246,985
; FILING DATE: 20-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-464-134-106

```

```

Query Match      8.4%; Score 54; DB 2; Length 108;
Best Local Similarity 38.2%; Pred. No. 33;
Matches 13; Conservative 6; Mismatches 13; Indels 2; Gaps 1;

```

```

Qy 68 DGHEVFYGMFDAGSTGTRVHVVFQFTRPPRPTPT 101
Db 18 EGHAV--GMLVSVLHSGGRVTAARTRPWTQVPT 49

```

## RESULT 8

```

US-08-461-361-106
; Sequence 106, Application US/08461361
; Patent No. 5856134
; GENERAL INFORMATION:
; APPLICANT: Kim, Jungsuh P.
; APPLICANT: Wages, John

```

```

; APPLICANT: Young, LaVonne M.
; APPLICANT: Fry, Kirk E.
; APPLICANT: Limen, Jeffrey M.
; TITLE OF INVENTION: Hepatitis G Virus and Molecular
; Cloning Thereof
; NUMBER OF SEQUENCES: 277
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave., Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,361
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/389,886
; FILING DATE: 15-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/357,509
; FILING DATE: 16-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,729
; FILING DATE: 26-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/344,271
; FILING DATE: 23-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/285,558
; FILING DATE: 03-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/285,543
; FILING DATE: 03-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/246,985
; FILING DATE: 20-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-461-361-106

```

```

Query Match      8.4%; Score 54; DB 2; Length 108;
Best Local Similarity 38.2%; Pred. No. 33;
Matches 13; Conservative 6; Mismatches 13; Indels 2; Gaps 1;

```

```

Qy 68 DGHEVFYGMFDAGSTGTRVHVVFQFTRPPRPTPT 101
Db 18 EGHAV--GMLVSVLHSGGRVTAARTRPWTQVPT 49

```

## RESULT 9

```

US-08-485-910-106
; Sequence 106, Application US/08485910
; Patent No. 5874563
; GENERAL INFORMATION:
; APPLICANT: Kim, Jungsuh P.

```

APPLICANT: Wages, John  
APPLICANT: Young, Lavonne M.  
APPLICANT: Fry, Kirk E.  
APPLICANT: Linnen, Jeffrey M.  
TITLE OF INVENTION: Hepatitis G Virus and Molecular  
TITLE OF INVENTION: Cloning thereof  
NUMBER OF SEQUENCES: 277  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: 350 Cambridge Ave., Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,910  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/389,886  
FILING DATE: 15-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/357,509  
FILING DATE: 16-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,729  
FILING DATE: 26-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/344,271  
FILING DATE: 23-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/285,558  
FILING DATE: 03-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/285,543  
FILING DATE: 03-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/246,985  
FILING DATE: 20-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Fabian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 106:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 108 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-485-910-106

Query Match 8.4%; Score 54; DB 2; Length 108;  
Best Local Similarity 38.2%; Pred. No. 33;  
Matches 13; Conservative 6; Mismatches 13; Indels 2; Gaps 1;

Qy 68 DGHEVYGIIMPDAGSTGTRVHVFOTRPPRETPT 101  
Db 18 EGHAV--GMLVSVLHSGGRVTAARTRPWTQVPT 49

RESULT 10  
PCT-US95-06266-89  
Sequence 89, Application PC/TUS9506266  
GENERAL INFORMATION:  
APPLICANT:

TITLE OF INVENTION: Detection of Viral Antigens Coded  
TITLE OF INVENTION: by Reverse Reading Frames  
NUMBER OF SEQUENCES: 157  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/06266  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/246,985  
FILING DATE: 20-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/285,561  
FILING DATE: 03-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,729  
FILING DATE: 26-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/344,271  
FILING DATE: 23-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/357,509  
FILING DATE: 16-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/389,886  
FILING DATE: 15-FEB-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Fabian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 4600-0202.41  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 89:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 108 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-06266-89

Query Match 8.4%; Score 54; DB 5; Length 108;  
Best Local Similarity 38.2%; Pred. No. 33;  
Matches 13; Conservative 6; Mismatches 13; Indels 2; Gaps 1;

Qy 68 DGHEVYGIIMPDAGSTGTRVHVFOTRPPRETPT 101  
Db 18 EGHAV--GMLVSVLHSGGRVTAARTRPWTQVPT 49

RESULT 11  
US-09-543-681A-6609  
Sequence 6609, Application US/09543681A  
Patent No. 6605709  
GENERAL INFORMATION:  
APPLICANT: GARY BRETON  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
FILE REFERENCE: 2709.1002-001  
CURRENT APPLICATION NUMBER: US/09/543,681A  
CURRENT FILING DATE: 2000-04-05  
PRIOR APPLICATION NUMBER: US 60/128,706

; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 6609  
; LENGTH: 109  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-6609

Query Match 8.3%; Score 53; DB 4; Length 109;  
Best Local Similarity 33.3%; Pred. No. 45;  
Matches 17; Conservative 7; Mismatches 19; Indels 8; Gaps 3;

QY 19 LGLCVGVFTYVAYIKW---HRATATQAFPSITRAAPGARW---GQOAHSP 62  
DB 43 LHLCTV-LFTVPEKKWKLKLLASQILSVPLCGDAGSSLVARRHP 92

## RESULT 12

US-09-489-039A-8242  
; Sequence 8242, Application US/09489039A  
; Patent No. 6610836

; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 8242  
; LENGTH: 113  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-8242

Query Match 8.0%; Score 51.5; DB 4; Length 113;  
Best Local Similarity 25.8%; Pred. No. 73;  
Matches 25; Conservative 10; Mismatches 35; Indels 27; Gaps 6;

QY 33 KHRATATQAFPSITRAAPG---ARWGOAHSPGTAADGHEVFFYGMFDAGSTGTRVHV 89  
DB 14 RWRSSAPG---RAGRGSGRRNRPFGDSHP--APSADYR-----ARSPARRSHA 60  
QY 90 FQFT--RPP-----RETPTLTHTETKAVKPGLS 115  
DB 61 GRARCRPPAADGHRRTIRDPPGPAAGCYHAAPPALS 97

## RESULT 13

US-09-489-039A-12049  
; Sequence 12049, Application US/09489039A  
; Patent No. 6610836

; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 12049  
; LENGTH: 101  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-12049

Query Match 8.0%; Score 51; DB 4; Length 101;  
Best Local Similarity 25.8%; Pred. No. 73;  
Matches 27; Conservative 7; Mismatches 38; Indels 34; Gaps 4;

QY 37 ATATQAFPSI-----TRAAPG-----ARWGOAHSPGLG-----TAADGHEV 72  
DB 6 AVAWSAFAAVDYRHYASAGSGSPAAGQGRQADAPAGQPVVAAVRAATQSDRTAAAA--- 62  
QY 73 FYGMFDAGSTGTRVHVFOFTRPPTPTLTHTETKAVKPGLSAYA 118  
DB 63 -----AQADGVSAADAQVDNPAERPPVVDSDGAPAAAGLPAWA 101

## RESULT 14

US-08-959-212-10  
; Sequence 10, Application US/08959212  
; Patent No. 6060274

; GENERAL INFORMATION:  
; APPLICANT: Bjornvad, Mads  
; APPLICANT: Schulein, Martin  
; APPLICANT: Jorgensen, Per  
; TITLE OF INVENTION: Extracellular Expression Of Cellulose  
; FILE REFERENCE: 4987.200-US  
; CURRENT APPLICATION NUMBER: US/08/959,212  
; CURRENT FILING DATE: 1997-10-28  
; EARLIER APPLICATION NUMBER: 1192/96  
; EARLIER FILING DATE: 1996-10-28  
; EARLIER APPLICATION NUMBER: 1426/96  
; EARLIER FILING DATE: 1996-12-13  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 10  
; LENGTH: 108  
; TYPE: PRT  
; ORGANISM: Humicola insolens  
US-08-959-212-10

Query Match 8.0%; Score 51; DB 3; Length 108;  
Best Local Similarity 27.7%; Pred. No. 80;  
Matches 23; Conservative 10; Mismatches 30; Indels 20; Gaps 4;

QY 22 CVGVFTYVAYIKWH-----RATATQAFPSITRAAPGARWGOAHSPGLGTAAD--- 68  
DB 23 CVSGYTCVLYNDWVYSCQCPQPTTLRTTTTTPGATSTTTSAPAA---TSITTPAGCTAERWA 78  
QY 69 --GHEVFG-IMFDAGSTGTRVH 88  
DB 79 QCGNGWMSGCTTCVAGSTCTKIN 101

## RESULT 15

US-09-621-976-4538  
; Sequence 4538, Application US/09621976  
; Patent No. 6639063

; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 4538  
; LENGTH: 90  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: 83  
; OTHER INFORMATION: Xaa = Ala, Pro  
; NAME/KEY: UNSURE  
; LOCATION: 15  
; OTHER INFORMATION: Xaa = Phe, Leu



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model1

Run on: June 8, 2004, 10:06:37 ; Search time 20 Seconds  
(without alignments)  
1548.684 Million cell updates/sec

Title: US-09-905-589A-2\_COPY\_135\_456  
Perfect score: 1659  
Sequence: 1 AKQIDFPDMKATPLVKAT.....ALGAIHFHYIDSLNRQSPAS 322

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 167602

Minimum DB seq length: 0  
Maximum DB seq length: 322

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	94	5.7	280	2	T28684
2	86	5.2	316	2	D87839
3	86	5.2	316	2	T21451
4	85.5	5.2	244	2	T36357
5	84.5	5.1	246	2	H39999
6	84	5.1	293	2	G72334
7	83	5.0	212	1	S04473
8	82	4.9	212	1	U60706
9	82	4.9	297	2	E36002
10	82	4.9	322	2	T27308
11	81.5	4.9	267	2	AH3234
12	81	4.9	223	2	B37068
13	80.5	4.9	306	1	J01395
14	79.5	4.8	274	1	SRBSD
15	79	4.8	322	2	T24948
16	78.5	4.7	245	2	D72574
17	77.5	4.7	314	2	H64531
18	77	4.6	285	2	AF0146
19	77	4.6	318	2	T50752
20	76.5	4.6	252	2	PC4259
21	75.5	4.6	272	2	S61888
22	75.5	4.6	320	2	E71139
23	75.5	4.6	322	2	S77066
24	75	4.5	250	2	D70787
25	75	4.5	297	2	F83491
26	74.5	4.5	163	2	G95338
27	74.5	4.5	222	2	G72232
28	74.5	4.5	223	2	B71499
29	74.5	4.5	240	2	AC3266

30	74.5	4.5	274	1	D70378	conserved hypothet
31	74.5	4.5	322	2	D97349	glycosyltransferas
32	74	4.5	128	2	T28433	selenophosphate sy
33	73.5	4.4	227	2	S28440	hypothetical prote
34	73.5	4.4	249	2	G75048	proliferating-cell
35	73.5	4.4	284	2	A83688	hypothetical prote
36	73.5	4.4	288	2	D75286	serine proteinase
37	73.5	4.4	301	2	JC2039	phosphoribosylamin
38	73.5	4.4	306	2	E72337	hypothetical prote
39	73	4.4	197	2	S28574	ig light chain - A
40	73	4.4	259	2	E95903	probable oxidoredu
41	73	4.4	275	2	T32813	hypothetical prote
42	73	4.4	281	2	D95395	probable Beta lact
43	73	4.4	290	2	D82545	succinyl-CoA synth
44	72.5	4.4	197	2	H90211	conserved hypothet
45	72.5	4.4	249	2	AD2827	short chain dehydr

## ALIGNMENTS

RESULT 1  
T28684  
hypothetical protein - Streptomyces coelicolor  
C/Species: Streptomyces coelicolor  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 20-Jun-2000  
C/Accession: T28684  
R:Parkhill, J.; Bentley, S.D.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, January 1999  
A/Reference number: Z20512  
A/Accession: T28684  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-280 <PAR>  
A/Cross-references: EMBL:AL023496; PIR:CA18901.1  
C/Superfamily: hypothetical protein YLR351c

Query Match 5.7%; Score 94; DB 2; Length 280;  
Best Local Similarity 31.1%; Pred. No. 1.5;  
Matches 37; Conservative 13; Mismatches 53; Indels 16; Gaps 5;  
QY 147 GVEGQPAKD---GKEIVSPCLSPFKGEV-----EHAETVYVSGQKAAS-----LHEL 193  
Db 136 GNSGWPVFTPAVGKGVYCYDRHPEEGRGALGLGAEIVFPASNSRLSGYIMQLBP 195  
QY 194 CAARVSEVLQNVHR--TEEVKHDVFPASYYIDLAG-VGLIDAKGSLVGPPEIA 249  
Db 196 AAAVNEVFGAIVGVGVELGDNDYFGTSYVDPEARFVGVAASDKETELVVRDLDVA 254

RESULT 2  
D87839  
protein F27D4.1 [imported] - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 27-Oct-2003  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A/Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog  
A/Reference number: A75000; MIMD:19969613; PMID:2851916  
A/Note: see website genome.wustl.edu/gsc/C\_elegans/ and www\_sanger.ac.uk/Projects/C\_ele  
A/Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A/Accession: D87839  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-316 <STO>  
A/Cross-references: GB:chr\_I; PIRN:CA01967.1; PIR:G3876390; GSPDB:GN00019; CESP:F27D4.1  
C/Genes: C\_elegans: F27D4.1  
A/Map position: 1  
C/Superfamily: electron transfer flavoprotein alpha subunit  
Query Match 5.2%; Score 86; DB 2; Length 316;



Best Local Similarity 23.9%; Pred. No. 9.1; Matches 67; Conservative 26; Mismatches 103; Indels 84; Gaps 13;  
Qy 8 DFWKATPLVKA-TAGLR-----LLPGEKAQKLLQKVEFKASPFLLVGDDCVSIWNG 59  
Db 14 DETKLAPITLNAITAAASKLGNEVSVLTGANATKVAEQVAKVNGKRVLVAQD-EKLNKN 72  
Qy 60 TDEGVSAWI-----TINELTGLSKTPGSSVGMGLDGGSGTQIAFLPRVEGTLOASPPGY 114  
Db 73 LPERVAPVILASQKQNF---TAITAGSAFGR-----GVIPRVAAKLDVSSISD 119  
Qy 115 LTALR---MFNRTYKLYS-----YSYLGGLMSARLAILGGVEGQPAKDG 156  
Db 120 VTEVHSDSPTRT--LYAGNAVKKVKTAPIKLLTFRGTSFEPKAGGSGGAVENAPSADI 177  
Qy 157 K-----ELVSPCLSPSPKGEWEHAEVTVRSVQKAAASLHELCAARVSEVLQNRVHRTTEV 212  
Db 178 KTDLSSEFLGQELSKSRPDLATAKV--VSGRGLKSGDN----- 215  
Qy 213 KHVDYAFSYYYDLA----AGVGLIDAEGGSLVVGDFEI 248  
Db 216 -----FKLIYDLADKLGAAGVGSRAAADVGVFNDMQV 248  
RESULT 3  
T21451  
hypothetical protein F27D4.1 - Caenorhabditis elegans (fragment)  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 27-Oct-2003  
C:Accession: T21451  
R:Wilkinson, J.  
submitted to the EMBL Data Library, September 1996  
A:Reference number: Z19424  
A:Accession: T21451  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-316 <WIL>  
A:Cross-references: EMBL:Z79695; PIDN:CAB01967.1; GSPDB:GN00019; CESP:F27D4.1  
A:Experimental source: clone F27D4  
C:Genetics:  
A:Map position: 1  
A:Gene: CESP:F27D4.1  
C:Superfamily: electron transfer flavoprotein alpha subunit  
Query Match 5.2%; Score 86; DB 2; Length 316;  
Best Local Similarity 23.9%; Pred. No. 9.1; Matches 67; Conservative 26; Mismatches 103; Indels 84; Gaps 13;  
Qy 8 DFWKATPLVKA-TAGLR-----LLPGEKAQKLLQKVEFKASPFLLVGDDCVSIWNG 59  
Db 14 DETKLAPITLNAITAAASKLGNEVSVLTGANATKVAEQVAKVNGKRVLVAQD-EKLNKN 72  
Qy 60 TDEGVSAWI-----TINELTGLSKTPGSSVGMGLDGGSGTQIAFLPRVEGTLOASPPGY 114  
Db 73 LPERVAPVILASQKQNF---TAITAGSAFGR-----GVIPRVAAKLDVSSISD 119  
Qy 115 LTALR---MFNRTYKLYS-----YSYLGGLMSARLAILGGVEGQPAKDG 156  
Db 120 VTEVHSDSPTRT--LYAGNAVKKVKTAPIKLLTFRGTSFEPKAGGSGGAVENAPSADI 177  
Qy 157 K-----ELVSPCLSPSPKGEWEHAEVTVRSVQKAAASLHELCAARVSEVLQNRVHRTTEV 212  
Db 178 KTDLSSEFLGQELSKSRPDLATAKV--VSGRGLKSGDN----- 215  
Qy 213 KHVDYAFSYYYDLA----AGVGLIDAEGGSLVVGDFEI 248  
Db 216 -----FKLIYDLADKLGAAGVGSRAAADVGVFNDMQV 248

RESULT 4  
T36357  
probable membrane protein - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T36357  
R:Oliver, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, May 1999  
A:Reference number: Z21575  
A:Accession: T36357  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-244 <OLI>  
A:Cross-references: EMBL:AL049841; PIDN:CAB42784.1; GSPDB:GN00070; SCOEDB:SCB9.38  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SCB9.38  
Query Match 5.2%; Score 85.5; DB 2; Length 244;  
Best Local Similarity 23.3%; Pred. No. 7.1; Matches 53; Conservative 23; Mismatches 92; Indels 59; Gaps 9;  
Qy 67 WITINFLTGLSKTPGSSVGMGLDGGSGTQIAFLPRVEGTLOASPPGY-----LTALRMF 121  
Db 20 WVTAGNTASTENTASAESTS---GAWRTGVA-----AGTASATAPGTDVRSQGVAFRRS 70  
Qy 122 NRTVKLYSYSLGLMSARLAIL-----GGVEGQPAKDGKELVSPCLSPS----- 167  
Db 71 RRTSRVVAALSLSGLVGA-LALTACTDGGSDGGSGEGGDKSSAFAFGASATDPGG 129  
Qy 168 -----PKGEW-----EHAETVTVRSVQKAA--ASLHELCAARVSEVLQNRVHR 208  
Db 130 DAGGSPSAGAAGELEGSWLATTDGQAVLMTGDKKALFATGTVCSGTTEETSGTTRIR 189  
Qy 209 -----TEEVKHVDYAFSYYYDLAAGVGLIDAEGGSLVVG 244  
Db 190 LKCADGSGADRATGKGVAGATSLTVAWEGALGKETVTRSEGGSLPPG 236  
RESULT 5  
H39999  
hypothetical protein 8 - Chlamydomonas reinhardtii  
C:Species: Chlamydomonas reinhardtii  
C:Date: 08-Nov-1991 #sequence\_revision 05-May-2000 #text\_change 09-Dec-2002  
A:Accession: H39999; S18149  
R:Lusher, M.E.; Gregory, J.; Storey, C.C.; Richmond, S.J.  
submitted to the EMBL Data Library, October 1991  
A:Description: Analysis of the complete nucleotide sequence of the plasmid pCPA1 isolate  
submitted to the Protein Sequence Database, October 1991  
A:Reference number: A39999  
A:Accession: H39999  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-246 <LUS>  
R:Lusher, M.E.; Gregory, J.; Storey, C.C.; Richmond, S.J.  
submitted to the EMBL Data Library, October 1991  
A:Description: Analysis of the complete nucleotide sequence of the plasmid pCPA1 isolate  
submitted to the Protein Sequence Database, October 1991  
A:Reference number: S18141  
A:Accession: S18149  
A:Molecule type: DNA  
A:Residues: 1-99 <LU2>  
A:Cross-references: EMBL:X62475; NID:G40608; PIDN:CAA44340.1; PID:G581000  
A:Accession: S18141  
A:Molecule type: DNA  
A:Residues: 100-246 <LU3>  
A:Cross-references: EMBL:X62475; NID:G40608; PIDN:CAA44332.1; PID:G40609  
C:Genetics:  
A:Start codon: GTG  
C:Superfamily: Virulence pgp6-D-related protein  
Query Match 5.1%; Score 84.5; DB 2; Length 246;  
Best Local Similarity 20.7%; Pred. No. 8.8; Matches 44; Conservative 40; Mismatches 84; Indels 45; Gaps 10;  
Qy 18 KATAGRLRLPGEKAQKLLQKVEFKASPFLLVGDDCVSI-----MNGTDEGVSAWITIN 71  
Db 62 RGLASIKLLTQ-----IKSIQKQHVLLIGEKIYKVRILKMNNSPDFTTFSWINLV 113  
Qy 72 FLTGLSKTPGSSVGMGLDGGSGTQIAFLP--RVEGTLOASPPGYLTALRMFNRTYKLYS 129

Db 114 FRT---KSAVNALGGYEL-----FSLDPKNTKSLFOSIP---YNTAYILASRK----- 157  
Qy 130 YSYLGLGMSARLAILIGVGQPAKDGKELVSPCLSPSFGKWEHAEVTVYRSGQRAAS 189  
Db 158 -----GSVXDKLKVLSIGLSNALADIVLNKFLPLKSSQTERC-VDFEEKKEVSEK 210  
Qy 190 LHEL-----CAARSEVTLQNRVHRT-BEYGHV 216  
Db 211 LIDILKIVSGGLSEYKNKLLHQLFEKTLKVD 243

## RESULT 6

G72334 [acyl-carrier-protein] S-malonyltransferase (EC 2.3.1.39) TM0798 [similarity] - Thermoc  
C/Species: Thermocoga maritima  
C/Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C/Accession: G72334  
R/Neison, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.V.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.  
C.M.  
Nature 399, 323-329, 1999  
A/Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A/Reference number: A72200; MUID:99287316; PMID:10360571  
A/Accession: G72334  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-293 <AB>  
A/Cross-references: GB:AA001747; GB:AA000512; NID:94981304; PIDD:AA035880.1; PID:9498132  
A/Experimental source: strain MSB  
C/Genetics:  
A/Gene: TM0798  
C/Superfamily: [acyl-carrier-protein] S-malonyltransferase; [acyl-carrier-protein] S-mal  
C/Keywords: acyltransferase; coenzyme A  
P/3-381/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>  
F/87/Active site: Ser (covalent substrate-binding) #status predicted  
F/193/Active site: His #status predicted

Query Match 5.1%; Score 84; DB 2; Length 293;  
Best Local Similarity 24.2%; Pred. No. 12;  
Matches 59; Conservative 32; Mismatches 67; Indels 86; Gaps 13;

Qy 25 LIPGKAAQ-----KLLQKVEYFKASPLVGDCCSINMGDEGV-----SAMI 68  
Db 5 VFPGGSGSYGKAMDPFVYESKELFEKSKVLGFDITENNGDETLKLTENAPSIYI 64  
Qy 69 TINFPTGSLKTPGGSSVGMGLDGGGSGTQIAFLPRVREGTLQASPGYVLTALMFRRTKLY 128  
Db 65 T-----SYIAFLEL-----EKRGILPDV---VAHSHLGEYITALAVAG-VYDFE 103  
Qy 129 SYSYV-----GLGMSARLAILIGVGQPAKDGKELVSPCLSPSFGKEM-- 172  
Db 104 TGLYLVRRKGEYMSKALEPFGKGTMAA-----VGLNIETIEEVN-----SIEGYVIA 151  
Qy 173 ---EHAETVYRSG-----QKRAASLHELCAARSEVTLQNRVHRT-----BEVX 213  
Db 152 NYNSHDQVV--ISGLKSEVERKAMETLKEKARRAVELNVSSPFHTPLFLEYAREMKKEBE 209  
Qy 214 HVDF 217  
Db 210 KYDF 213

RESULT 7  
S04473 nitrile hydratase (EC 4.2.1.84) beta chain [validated] - Rhodococcus sp.  
C/Species: Rhodococcus sp.  
C/Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 27-Oct-2003  
C/Accession: S04473; S02071; S16072; S54106; S62096  
R/Durak, O.; Nishiyama, M.; Horinouchi, S.; Beppu, T.  
Eur. J. Biochem. 181, 563-570, 1999  
A/Title: Primary structure of nitrile hydratase deduced from the nucleotide sequence of  
A/Reference number: S04471; MUID:89276538; PMID:2659343

A/Accession: S04473  
A/Molecule type: DNA  
A/Residues: 1-212 <IK>  
A/Cross-references: EMBL:X14668; NID:946429; PIDD:CA432798.1; PID:946431  
A/Note: part of this sequence, including the amino end of the mature protein, was confi  
R/Endo, T.; Watanabe, I.  
FEBS Lett. 243, 61-64, 1989  
A/Title: Nitrile hydratase of Rhodococcus sp. N-774. Purification and amino acid sequenc  
A/Reference number: S02070; MUID:89153549; PMID:2920826  
A/Accession: S02071  
A/Molecule type: protein  
A/Residues: 1-19 <END>  
R/Hashimoto, Y.; Nishiyama, M.; Ikehata, O.; Horinouchi, S.; Beppu, T.  
Biochim. Biophys. Acta 1088, 225-233, 1991  
A/Title: Cloning and characterization of an amide gene from Rhodococcus species N-774  
A/Reference number: S15070; MUID:91159474; PMID:2001397  
A/Accession: S15072  
A/Molecule type: DNA  
A/Residues: 1-12/179-212 <HAS>  
A/Cross-references: EMBL:X54074  
A/Experimental source: strain N-774  
R/Bigey, F.; Chebrou, H.; Arnaut, A.; Galay, P.  
submitted to the EMBL Data Library, March 1995  
A/Description: Cloning, sequencing of the modified nitrile hydratase gene from mutant st  
A/Reference number: S54104  
A/Accession: S54106  
A/Molecule type: DNA  
A/Residues: 1-39, 'V', 41-212 <BIG>  
A/Cross-references: EMBL:Z48769; NID:9769823; PIDD:CA88686.1; PID:9769826  
R/Huang, W.; Schneider, G.; Lindqvist, Y.  
submitted to the Brookhaven Protein Data Bank, April 1997  
A/Reference number: A73039; PDB:1AHO  
A/Contents: annotation, X-ray crystallography, 2.65 angstroms  
R/Nagashima, S.; Nakasako, M.; Dohmae, N.; Tsujimura, M.; Takio, K.; Odaka, M.; Yohada,  
Nat. Struct. Biol. 5, 347-351, 1998  
A/Title: Novel non-heme iron center of nitrile hydratase with a claw setting of oxygen e  
A/Reference number: A58907; MUID:98246406; PMID:9586994  
A/Contents: annotation, X-ray crystallography, 1.7 angstroms; mass spectroscopic identifi  
C/Function:  
A/Description: catalyzes the reaction of one molecule of water and an aliphatic nitrile  
C/Superfamily: nitrile hydratase; beta subunit  
C/Keywords: carbon-oxygen lyase; hydro-lyase; metalloprotein  
P/-212/Product: nitrile hydratase beta chain #status experimental <MAT>  
P/56,111/Binding site: nitrosyl iron (Arg) (shared with alpha chain) #status experimente

Query Match 5.0%; Score 83; DB 1; Length 212;  
Best Local Similarity 29.0%; Pred. No. 9.7;  
Matches 38; Conservative 17; Mismatches 46; Indels 30; Gaps 8;

Qy 145 LGVVEGPAKDGK-ELVSPCLSPSFGKWEHAEVTVYRSGQRAASLHELCAARSEVTL 202  
Db 7 LAGVQG---FGKVPRTVADIGPTFAHEWELPSLNRAG-----VAELGAFVDEY- 55  
Qy 203 QNRVHRTBEYKRVDFAFASYVD-LAAGVGLIDAEKG-----GSLVGDPELIARVYC 254  
Db 56 -RYVERMEPRH---YMTPIYERYVIGVATLMVERGILTDQLSLAGGPPLS----- 106  
Qy 255 RILETQPOSSP 265  
Db 107 RPSESEGRAP 117

## RESULT 8

JN0706 nitrile hydratase (EC 4.2.1.84) beta chain - Rhodococcus erythropolis  
C/Species: Rhodococcus erythropolis  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 27-Oct-2003  
C/Accession: JN0706  
R/Durak, O.; Nishiyama, M.; Horinouchi, S.; Beppu, T.  
Biochem. Biotechnol. Biochem. 57, 1323-1328, 1993  
A/Title: Characterization of nitrile hydratase genes cloned by DNA screening from Rhodoc  
A/Reference number: P06570; MUID:93379358; PMID:7764017  
A/Accession: JN0706



Db 54 -----VCTTGAAGCGGKSTLLR-----IVAGLEPQFGGTV 83  
 QY 153 AKDGKEIVSPCLSPSPFKGEMHAEVTVKVSOKAAASLHELCAARVSEVLQNRVHRTVEV 212  
 Db 84 LLGGKPLGPGDLGRIVPQ-DHRLVPMTVTANLAFSLHRLPKAEQRVV-----TEKL 136  
 QY 213 KHVDFYAF--SYIYDLAAGVGLIDAEKGGSLVGDFFELIAKYVCTETPQQ---SSPF 266  
 Db 137 KLVGLEGGRSYPHQLSGMAQRYA-----IARALAPQELLILDEPF 179  
 QY 267 SCMD-LTVYSLLTLEFGFPRSKVLTETKIDNVER 300  
 Db 180 GADLALTRLOM-----QDEVLRI-RHTDNLTT 205

## RESULT 12

B97068

cobalamin biosynthesis protein CblM [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001

C:Accession: B97068

R:Noiling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: B97068

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-223 &lt;KUR&gt;

A:Cross-references: GB:AE001437; PIDN:AAK79333.1; PID:G15024300; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

A:Genetics:

A:Gene: CAC1365

C:Superfamily: cobalamin biosynthesis protein M

Query Match 4.9%; Score 81; DB 2; Length 223;  
 Best Local Similarity 23.4%; Pred. No. 16;  
 Matches 43; Conservative 28; Mismatches 71; Indels 42; Gaps 7;  
 QY 17 LKATAGRLRLPBGKAKLQKYEVPKASPF---VGDGCVSINMGDEGVSAITTNF 72  
 Db 16 LVAAASLFFVPG--VAEIKRSEKENVLYKPFISMVGVGVFVISCW-----IYVP- 63  
 QY 73 LKSLKTPGGSSVGMDLGGSGTQIAFLPRVEGTLQASPPGYLTALRFNFTYKLYSXY 132  
 Db 64 VTGSGCHPGCTPLAAILIGPLATSV-----ITIALFPQALFLGHGCI 106  
 QY 133 LGLGMSARLAILLGVGEGP---AKDGKEIVSPCLSPSPFKGEMHAEVTVKVSOKAA 187  
 Db 107 TTTIGANNISMGIAISGYAFKFLFRKFGSVYLAAGVAGVGDV---VYVMSALIELA 162  
 QY 188 ASLIL 191  
 Db 163 VSLH 166

## RESULT 13

J01395

phosphoribosylaminoimidazoleleucylsuccinocarboxamide synthase (EC 6.3.2.6) - yeast (Saccharomy

N:Alternate names: protein YAR015W; SAICAR synthetase

C:Species: Saccharomyces cerevisiae

C:Date: 17-Jul-1992 #sequence\_revision 03-Nov-1995 #text\_change 05-Nov-1999

C:Accession: S20122; S40905; S25679; J01395

R:Davies, C.J.; Hutchison III, C.A.

Nucleic Acids Res. 19, 5731-5738, 1991

A:Title: A directed DNA sequencing strategy based upon Tn3 transposon mutagenesis: appl

A:Reference number: S20121; MUID:92051323; PMID:1658741

A:Accession: S20122

A:Molecule type: DNA

A:Residues: 1-306 &lt;DAV&gt;

A:Cross-references: EMBL:M67445; NID:g170997; PIDN:AAA34398.1; PID:g170999

R:Clark, M.W.; Keng, T.; Storms, R.K.; Zhong, W.; Fortin, N.; Zeng, B.; Delaney, S.; Que  
 submitted to the EMBL Data Library, November 1993

A:Description: Sequencing of chromosome I of Saccharomyces cerevisiae: Analysis of the

A:Reference number: S40891

A:Accession: S40905

A:Molecule type: DNA

A:Residues: 1-306 &lt;CLA&gt;

A:Cross-references: EMBL:122015; NID:g1339990; PIDN:AA04963.1; PID:g349755; GSPDB:GN00

R:Schweitzer, B.; Phillipsen, P.

Mol. Gen. Genet. 234, 164-167, 1992

A:Title: NPK1, a nonessential protein kinase gene in Saccharomyces cerevisiae with simi

A:Reference number: S23580; MUID:92357012; PMID:1495480

A:Accession: S25679

A:Molecule type: DNA

A:Residues: 285-306 &lt;SCH&gt;

A:Cross-references: EMBL:X60549; NID:g298024; PIDN:CAA43043.1; PID:g4043

R:Myasnikov, A.N.; Sasnauskas, K.V.; Janulaitis, A.A.; Smirnov, M.N.

Gene 109, 143-147, 1991

A:Title: The Saccharomyces cerevisiae ADE1 gene: structure, overexpression and possible

A:Reference number: J01395; MUID:92097943; PMID:1756975

A:Accession: J01395

A:Molecule type: DNA

A:Residues: 1-184; G, 186-306 &lt;MTA&gt;

A:Cross-references: GB:M61209; NID:g170995; PIDN:AAA34396.1; PID:g170996

A:Experimental source: strain FH4C

A:Note: the authors translated the codon TAC for residue 98 as Asp

C:Genetics:

A:Gene: SGD:ADE1; MIPS:YAR015W

A:Cross-references: SGD:S0000070; MIPS:YAR015W

A:Map position: 1R

C:Function:

A:Description: catalyzes formation of N-succinyl-5-aminoimidazole-4-carboxamide ribotide

A:Pathway: purine biosynthesis

C:Superfamily: phosphoribosylaminoimidazoleleucylsuccinocarboxamide synthase

C:Keywords: ligase; purine nucleotide biosynthesis

Query Match 4.9%; Score 80.5; DB 1; Length 306;  
 Best Local Similarity 21.8%; Pred. No. 26;  
 Matches 58; Conservative 40; Mismatches 91; Indels 77; Gaps 16;  
 QY 37 KXKEVFPASP---FLVGDGCVS---INMGV-----SAWITINFLGSLKTPG 81  
 Db 19 KVRDIIEVDAGTLFLPAITRISAYDIVIMENSIPKILITKSEFN--FKFLSDVYN-- 74  
 QY 82 GSSVGMDLGGSGTQIAFLPRVEGTLQASPPGYLTALRFNFTYKLYSXYGL----- 135  
 Db 75 ----HLVDIAPGKTITFDYLP-----AKLSPEKTKTLE--DRSLVHKKILPLEVTVRG 123  
 QY 136 ----GLMSARLAILIGYVG---QPADGKEIVSPCLSPSPFKGE-WHAETVTVKVSOKAA 187  
 Db 124 YITGSAKVEYKGTGATGKQPGQLKESQFPPEPITPTPKAEGGHDE--NISPAQAA 180  
 QY 188 ASLHELCAARVSEVLQNRVHRTVEVHVDFYASYYUDDLAAGVGLIDA-----EKGS 240  
 Db 181 ELVGEDLSRRVAFIA-----VLYSKCKRYAKKXGILLIDTKEFEIDKTNF 228  
 QY 241 LVVGDFFIAKYVCTETPQSSPF 266  
 Db 229 IIVVD-EVLT-----PDSSRF 243

## RESULT 14

SUBSD

subtilisin (EC 3.4.21.62) DY - Bacillus subtilis (strain DY)

N:Alternate names: alkaline serine proteinase

C:Species: Bacillus subtilis

A:Variety: strain DY

C:Date: 20-Sep-1984 #sequence\_revision 20-Sep-1984 #text\_change 02-Jul-1998

C:Accession: A00969; S02492

R:Nedkov, P.; Oberthur, W.; Braunitzer, G.

Biochem. Biophys. Res. Commun. 100, 421-430, 1985

A:Title: Determination of the complete amino-acid sequence of subtilisin DY and its com

A:Reference number: A00969; MUID:85279896; PMID:3927935

Search completed: June 8, 2004, 10:10:38  
Job time : 22 secs

```

RESULT 15
T24948
hypothetical protein T16A9.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T24948
R:McMurray, A.
submitted to the EMBL Data Library, July 1996
A:Reference number: Z19959
A:Accession: T24948
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-322 <WIL>
A:Cross-references: EMBL:Z77135; PIDN:CA80878.1; GSPDB:GN00023; CESP:T16A9.5
A:Experimental source: clone T16A9
C:Genetics:
A:Gene: CESP:T16A9.5
A:Map position: 5
A:Introns: 71/1; 291/3
C:Superfamily: Caenorhabditis elegans hypothetical protein V69E1A.1

Query Match      4.8%; Score 79; DB 2; Length 322;
Best Local Similarity 25.1%; Pred. No. 38;
Matches 53; Conservative 30; Mismatches 78; Indels 50; Gaps 10
QV 38 VKEVFK-ASPFVLGVDCVSINNGTDEGVSAWITNPLTGLKTPGGSSV-----GM 87

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 10:02:51 ; Search time 18 Seconds  
(without alignments)  
931.477 Million cell updates/sec

Title: US-09-905-589a-2\_COPY\_135\_456  
Perfect score: 1659  
Sequence: 1 AKODIPDFPKATPLVLKAT.....ALGAFHYIDSLNRKSPAS 322

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues  
Total number of hits satisfying chosen parameters: 76390

Minimum DB seq length: 0  
Maximum DB seq length: 322

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	84.5	5.1	246	1	GP6D_CHLPS
2	83	5.0	212	1	NPAB_RHOER
3	82.5	5.0	192	1	PM11_HUMAN
4	81.5	4.9	312	1	GBLP_LEICH
5	81.5	4.9	312	1	GBLP_LEIWA
6	79.5	4.8	274	1	SUBD_BACLI
7	78	4.7	287	1	YQ73_VIBPA
8	77	4.6	315	1	ENGIC_PORGI
9	77	4.6	318	1	BCHC_RHOSE
10	75.5	4.6	286	1	CLN8_HUMAN
11	75.5	4.6	322	1	Y712_SYNY3
12	75	4.5	250	1	YMI9_MYCTU
13	74.5	4.5	274	1	PPNK_AQUAE
14	73.5	4.4	227	1	YFVA_METTF
15	73.5	4.4	249	1	PCNA_PYRAB
16	73.5	4.4	301	1	PUR7_PICTA
17	73.5	4.4	305	1	SYGA_STRPY
18	73.5	4.4	306	1	PUR7_YEAST
19	73.5	4.4	308	1	SYGA_STRPY
20	73	4.4	303	1	PUR7_PICAN
21	72.5	4.4	266	1	SUBH_RHIME
22	72.5	4.4	305	1	SYGA_STRPY
23	72	4.3	254	1	ADH_DROLE
24	72	4.3	276	1	YGRF_ECOLI
25	72	4.3	316	1	LDH_BORBU
26	71.5	4.3	293	1	ALFI_PORGI
27	71	4.3	163	1	DEF3_SHEON
28	70.5	4.2	292	1	YG38_PYRPU
29	70.5	4.2	125	1	PHPI_HUMAN
30	70.5	4.2	275	1	KPUI_BACAN
31	70	4.2	284	1	TPM2_CHICK
32	70	4.2	183	1	OLEC_BRANA
33	70	4.2	258	1	FABI_ANASP

## ALIGNMENTS

RESULT 1	GP6D_CHLPS	STANDARD;	PRT;	246 AA.
ID	GP6D_CHLPS			
AC	Q46264; Q46256;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Virulence plasmid protein GP6-D.			
OS	Chlamydia psittaci (Chlamydia psittaci).			
OG	Plasmid pCPA1.			
OC	Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.			
OX	NCBI_TaxID=83554;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=N352;			
RX	MEDLINE=97346036; PubMed=9202459;			
RA	Thomas N.S., Lusher M., Storey C.C., Clarke I.N.;			
RT	"Plasmid diversity in Chlamydia."			
RL	Microbiology 143:1847-1854(1997).			
CC	-1- SIMILARITY: BELONGS TO THE UPF0137 (GP6-D) FAMILY.			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL collabora-			
CC	tion the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see <a href="http://www.isb.ch/announce/">http://www.isb.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb.ch">license@isb.ch</a> ).			
CC	-----			
DR	EMBL; X62475; CAA44340.1; -.			
DR	EMBL; X62475; CAA44332.1; -.			
DR	PIR; H39999; H39999.			
DR	InterPro; IPR003350; UPF0137.			
DR	PIfam; PF03677; UPF0137; 1.			
KW	Plasmid.			
SQ	SEQUENCE 246 AA; 28110 MW; 78366B3FA4152DFP CRC64;			
Query Match	5.1%; Score 84.5; DB 1; Length 246;			
Best local similarity	20.7%; Pred. No. 5.7; Mismatches 84; Indels 45; Gaps 10;			
Matches 44;	Conservative 40; Mismatches 84; Indels 45; Gaps 10;			
DB	18 KATAGIRLIPBEKAKLQKYKVEFKASPPFVGDVSI-----MNGTDEGVSAMITIN 71			
DB	62 RLAIASIKILTQ-----IKSIQKQVLLIGEKIVREILKNMNSPDTTFSWIMLV 113			
QY	72 PLTSGIKTPGGSVGMULDGGSGTQIARLP--RVEGTQASPPVYTLTBMFNRTYKYS 129			
DB	114 FRT---KSAIYNAUGTYEL-----FSLPDKNTISLFQSTP--YKTAYILASKR----- 157			
QY	130 YSYLGLGMSARLILGVEGQPAKDGKELVSPCLSPFKEMWHAETVYVSGQKAAS 189			
DB	158 -----GSVKOKLKVIGKISGNSNALIDVANKFLPPLKSSQTERC--VDPEKKKVESEK 210			
QY	190 LHEL-----CAARSEVQGNRVHRT--EEVKHYD 216			
DB	211 LIDILKIVSGILESEYKKNLHQLFEXTLRVD 243			

RESULT 2  
NHAB\_RHOER STANDARD; PRT; 212 AA.

AC P13449; Q59789;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Nitrile hydratase subunit beta (EC 4.2.1.84) (Nitrilase) (NHase).  
GN NTHB OR NHA2.  
OS Rhodococcus erythropolis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Nocardiaceae; Rhodococcus.  
OX NCBI\_TaxID=1833;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=N-774;  
RX MEDLINE=89276338; PubMed=2659343;  
RA Ikemata O., Nishiyama M., Horinouchi S., Beppu T.;  
RT "Primary structure of nitrile hydratase deduced from the nucleotide  
RT sequence of a Rhodococcus species and its expression in Escherichia  
RT coli.";  
RL Eur. J. Biochem. 181:563-570(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=N-774;  
RX MEDLINE=91159474; PubMed=2001397;  
RA Hashimoto Y., Nishiyama M., Ikemata O., Horinouchi S., Beppu T.;  
RT "Cloning and characterization of an amidase gene from Rhodococcus  
RT species N-774 and its expression in Escherichia coli.";  
RL Biochim. Biophys. Acta 1088:225-233(1991).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ACV2;  
RA Bigey F., Chabrou H., Arnaud A., Galzy P.;  
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=N-771;  
RA Nojiri M., Yoshida M., Odaka M., Matsushita Y., Tsujimura M.,  
RA Yoshida T., Takio K., Endo I.;  
RT "Functional expression of nitrile hydratase in Escherichia coli.";  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE OF 1-19.  
RC STRAIN=N-774;  
RX MEDLINE=89153549; PubMed=2920826;  
RA Endo T., Watanabe I.;  
RT "Nitrile hydratase of Rhodococcus sp. N-774. Purification and amino  
RT acid sequences.";  
RL FEBS Lett. 243:61-64(1989).  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY (2.65 ANGSTROMS).  
RC STRAIN=Brevibacterium sp. R312;  
RX MEDLINE=97341231; PubMed=9195885;  
RA Huang W., Jia J., Cummings J., Nelson M., Schneider G., Lindqvist Y.;  
RT "Crystal structure of nitrile hydratase reveals a novel iron centre  
RT in a novel fold.";  
RL Structure 5:691-699(1997).  
RN [7]  
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS), AND MASS SPECTROMETRY.  
RX MEDLINE=98246406; PubMed=9586994;  
RA Nagashima S., Nakasako M., Dohmae N., Tsujimura M., Takio K.,  
RA Odaka M., Yoshida M., Kamiya N., Endo I.;  
RT "Novel non-heme iron center of nitrile hydratase with a claw setting  
RT of oxygen atoms.";  
RL Nat. Struct. Biol. 5:347-351(1998).  
CC -1- FUNCTION: NHASE CATALYZES THE HYDRATION OF VARIOUS NITRILE  
CC COMPOUNDS TO THE CORRESPONDING AMIDES. INDUSTRIAL PRODUCTION OF  
CC ACRYLAMIDE IS NOW BEING DEVELOPED USING SOME OF THESE ENZYMES.  
CC -1- CATALYTIC ACTIVITY: An aliphatic amide = a nitrile + H(2)O.  
CC -1- SUBUNIT: Heterodimer of an alpha and a beta chain.

CC -1- SIMILARITY: Belongs to the nitrile hydratase subunit beta family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X54074; CAA38011.1; -;  
DR EMBL; X14668; CAA32798.1; -;  
DR EMBL; Z48769; CAA86686.1; -;  
DR EMBL; AB016078; BAA36598.1; -;  
DR PDB; 1AHJ; 08-APR-98.  
DR PDB; 2AHJ; 16-FEB-99.  
DR InterPro; IPR008990; E transp acc.  
DR InterPro; IPR003168; NHase\_beta.  
DR Pfam; PF02211; NHase\_beta\_1.  
DR PROSITE; PS001427; NHase\_beta; 1.  
KW Lyase; 3D-structure.  
FT VARIANT 40 5 M -> V (IN STRAIN ACV2).  
FT TURN 4 5  
FT TURN 8 9  
FT TURN 20 21  
FT TURN 31 34  
FT TURN 35 45  
FT TURN 46 46  
FT TURN 52 60  
FT TURN 61 61  
FT TURN 64 69  
FT HELIX 72 85  
FT TURN 87 88  
FT HELIX 92 99  
FT TURN 100 100  
FT TURN 125 126  
FT STRAND 128 131  
FT HELIX 144 146  
FT TURN 147 148  
FT STRAND 150 156  
FT STRAND 161 161  
FT HELIX 164 167  
FT TURN 168 170  
FT STRAND 177 177  
FT STRAND 179 185  
FT HELIX 186 190  
FT TURN 191 192  
FT STRAND 198 204  
FT HELIX 205 207  
FT STRAND 208 210  
SQ SEQUENCE 212 AA; 23487 MW; A0401CA4FC1C2CBE CRC64;  
Query Match 5.0%; Score 83; DB 1; Length 212;  
Best Local Similarity 29.0%; Pred. No. 6.4; Indels 30; Gaps 8;  
Matches 38; Conservative 17; Mismatches 46;  
Qy 145 LGVVEGQPAKDGK--ELVSPCLSPSPFKGEWEHAEVTVYRSGQKAAASLHELCAARVSEVL 202  
Db 7 LAGVQG---FGKVPHTVNADIGTPTFAEWEHLPSYLMFAG-----VAELGAFSVDEV- 55  
Qy 203 QNEVHRTVEVKHVDYFAFSVYTD-LAAGVGLIDAEG-----GSLVVGDEIARVYC 254  
Db 56 -RVVVERMEPRH---YMMTPFYERYVIGVATLMVKEGILTQDELESLAGPPPLS----- 106  
Qy 255 RTLETQPSQSP 265  
Db 107 RPSESEGRFAP 117  
RESULT 3  
RM11 HUMAN  
ID RM11 HUMAN STANDARD; PRT; 192 AA.  
AC Q9Y3B7; Q96Q73;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE 60S ribosomal protein L11, mitochondrial precursor (L11mt) (CGI-113).

GN MRPL11.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20272150; PubMed=10810093;

RA Lai C.-H., Chou C.-Y., Ch'ang L.-Y., Liu C.-S., Lin W.-C.;

RT "Identification of novel human genes evolutionarily conserved in Caenorhabditis elegans by comparative proteomics."

RL Genome Res. 10:703-713(2000).

RL [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=21293042; PubMed=11279069;

RA Suzuki T., Terasaki M., Takemoto-Hori C., Hanada T., Ueda T., Wada A., Watanabe K.;

RT "Structural compensation for the deficit of rRNA with proteins in the mammalian mitochondrial ribosome. Systematic analysis of protein components of the large ribosomal subunit from mammalian mitochondria."

RL J. Biol. Chem. 276:21724-21736(2001).

RN [3]

RP SEQUENCE FROM N.A.

RX TISSUE=Lymph;

RX MEDLINE=22385257; PubMed=12477932;

RA Strusberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buerger K.H., Scheffer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stjepeton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E., Brownstein M.D., Urdin T.B., Toshiyuki S., Carrinchi P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Morley K.C., Hale S.N., Garcia A.M., Gay L.J., Halik S.W., Villalón D.K., Murry D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalhus D.E., Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RL [4]

RP SEQUENCE OF 81-141 FROM N.A.

RX MEDLINE=21429115; PubMed=11543634;

RA Kameuchi N., Suzuki T., Uechi T., Magoori M., Kunba M., Higa S., Watanabe K., Tanaka T.;

RT "The human mitochondrial ribosomal protein genes: mapping of 54 genes to the chromosomes and implications for human disorders."

RL Genomics 77:65-70(2001).

RL -1- SUBCELLULAR LOCATION: Mitochondrial.

CC -1- SIMILARITY: Belongs to the L11p family of ribosomal proteins.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL, AF151871; AAD34108.1; -

DR EMBL, AB046638; BAB40843.1; -

DR EMBL, BC005002; AAH05002.1; -

DR EMBL, AB051338; BAB54928.1; -

DR HSP, P29395; 1MWS.

DR Genew; HGNC:14042; MRPL11.

DR InterPro; IPRO00911; Ribosomal L11.

DR InterPro; IPRO06519; Ribosomal L11bac.

DR Pfam; PF00298; Ribosomal L11; 1.

DR Pfam; PF03946; Ribosomal L11N; 1.

DR Pfam; PF001367; Ribosomal L11; 1.

DR Pfam; PF00649; RLI1; 1.

DR SMART; SM00649; RLI1; 1.

DR TIGRFAMs; TIGR01632; L11\_bact; 1. FALSE NEG.

DR PROSITE; PS00359; RIBOSOMAL\_L11; 1.

DR Ribosomal protein; Mitochondrion; Transit peptide.

FT TRANSIT 1 ?

FT CHAIN ? 192

FT 60S RIBOSOMAL PROTEIN L11.

SO SEQUENCE 192 AA; 20683 MW; 92506A2004BF278 CRC64;

Query Match 5.0%; Score 82.5; DB 1; Length 192;

Best Local Similarity 22.7%; Pred. No. 6.2;

Matches 42; Conservative 28; Mismatches 76; Indels 39; Gaps 8;

QY 131 SYIGLGMSRLILGVEQPAKDKELVSPCLSFSGEMHAEVTVYSGCKAAASL 190

DB 2 SKLGRARGLRKEVGVIRAIYRAGLAMPPLGP-----VLGQR-GVSI 46

QY 191 HELCA--ARVSEVLQN-----RHRTSEKVDFAFYDYDLAAGVLIDAEKG 238

DB 47 NQCKEENERTKDIKIGIPPTKILYKPRTEPIK-IGQPTVYFLKAAAGI-----EKG 100

QY 239 ---GSLVGDPEIATKYCRITLPTOPQSSPSCMDLYVSLIQFGPPRSKYLKTRK 294

DB 101 ARQTKVEKVAQ--LVTLKHVEIRAIKQDEAFALQVPLSSVRSIIIGNARSIGIRVXD 158

QY 295 IDNVE 299

DB 159 LSSEB 163

RESULT 4

GBLP LEICH STANDARD; PRT; 312 AA.

AC 027434;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Granule nucleotide-binding protein beta subunit-like protein (Antigen DE LACK) (LIP36) (P36LR).

OS Leishmania chagasi, and

OC Leishmania infantum.

OC Eukaryota; Eumetazoa; Kinetoplastida; Trypanosomatidae; Leishmania.

OX NCBI\_TaxID=44271, 5671;

RN [1]

RP SEQUENCE FROM N.A.

RX SPECIES=L.chagasi;

RX MEDLINE=95242101; PubMed=7725103;

RA Mougneau E., Altare F., Makil A.E., Zheng S., Coppola T., Wang Z.E., Waldmann R., Locksley R.M., Glatchenhaus N.;

RT "Expression cloning of a protective Leishmania antigen."

RL Science 268:563-566(1995).

RL [2]

RP SEQUENCE FROM N.A.

RX SPECIES=L.infantum; STRAIN=PB75;

RA Gonzalez-Aseguinolaza G., Larraga V.;

RT Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.

RL -1- DEVELOPMENTAL STAGE: EXPRESSED IN BOTH STAGES OF THE PARASITE LIFE CYCLE.

CC -1- SIMILARITY: Contains 7 WD repeats.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).



```
CC ENBL; U27569; AAA57576.1; -
DR ENBL; U49695; AAA91208.1; -
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 7.
DR PRINTS; PR00320; GPROTEINERPT.
DR PRODOM; PD000018; WD40; 4.
DR SMART; SM00320; WD40; 6.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS00682; WD_REPEATS_2; 5.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat; Antigen.
FT REPEAT 9 42 WD 1.
FT REPEAT 63 93 WD 2.
FT REPEAT 105 135 WD 3.
FT REPEAT 148 180 WD 4.
FT REPEAT 192 222 WD 5.
FT REPEAT 233 262 WD 6.
FT REPEAT 279 307 WD 7.
SQ SEQUENCE 312 AA; 34372 MW; 778321761BBCCD99 CRC64;

Query Match 4.9%; Score 81.5; DB 1; Length 312;
Best Local Similarity 23.1%; Pred.No.14;
Matches 36; Conservative 19; Mismatches 48; Indels 53; Gaps 7;

QY 53 CYSIMNGTDEGVA-WITINFLTGLSGKTPGSSVGMGLDLGGSTQIAFLPRVEGTLOA-- 109
D 70 CVSLAHATDYALTASW-
QY 110 SPGYL-----TALRMFNRTYKLYSYVLGGLMSARLAILGGVGPQAKDGKE--LV 160
D 116 SPDRLLIVSAGRDNRVWVN-----VAGECMHEFLRDGHEDWVS 154

QY 161 SPCLSPSFK-----GEWEHAEVTVRVSGQKAAASL 190
D 155 SICFSPSLEHPVIVSGSDNTIKVNVNGKCCRTL 190

RESULT 5
GBLP LEIMA STANDARD; PRT; 312 AA.
AC Q25306;
DT 15-JUL-1998 (Rel. 36, Created)
DT 18-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Guanine nucleotide-binding protein beta subunit-like protein (Antigen LACK).
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95242101; PubMed=7725103;
RA Mougneau E., Altare F., Wakil A.E., Zheng S., Coppola T., Wang Z.E., Waldmann R., Locksley R.M., Glaichenhaus N.;
RT "Expression cloning of a protective Leishmania antigen.";
RL Science 268:563-566(1995).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN BOTH STAGES OF THE PARASITE LIFE CYCLE.
CC -1- SIMILARITY: Contains 7 WD repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC ENBL; U27568; AAA97577.1; -
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 7.
DR PRINTS; PR00320; GPROTEINERPT.
```

```
DR PRODOM; PD000018; WD40; 4.
DR SMART; SM00320; WD40; 6.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS00682; WD_REPEATS_2; 5.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat; Antigen.
FT REPEAT 9 42 WD 1.
FT REPEAT 63 93 WD 2.
FT REPEAT 105 135 WD 3.
FT REPEAT 148 180 WD 4.
FT REPEAT 192 222 WD 5.
FT REPEAT 233 262 WD 6.
FT REPEAT 279 307 WD 7.
SQ SEQUENCE 312 AA; 34413 MW; B094B566ASACCID3 CRC64;

Query Match 4.9%; Score 81.5; DB 1; Length 312;
Best Local Similarity 23.1%; Pred.No.14;
Matches 36; Conservative 19; Mismatches 48; Indels 53; Gaps 7;

QY 53 CYSIMNGTDEGVA-WITINFLTGLSGKTPGSSVGMGLDLGGSTQIAFLPRVEGTLOA-- 109
D 70 CVSLAHATDYALTASW-
QY 110 SPGYL-----TALRMFNRTYKLYSYVLGGLMSARLAILGGVGPQAKDGKE--LV 160
D 116 SPDRLLIVSAGRDNRVWVN-----VAGECMHEFLRDGHEDWVS 154

QY 161 SPCLSPSFK-----GEWEHAEVTVRVSGQKAAASL 190
D 155 SICFSPSLEHPVIVSGSDNTIKVNVNGKCCRTL 190

RESULT 6
SUBD BACLI STANDARD; PRT; 274 AA.
AC P00781;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Subtilisin DY (EC 3.4.21.62).
GN APR.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE.
RC STRAIN=DY;
RX MEDLINE=84110064; PubMed=6420308;
RA Nedkov P., Oberthur W., Braunitz G.;
RT "Primary structure of subtilisin DY.";
RL Hoppe-Seyler's Z. Physiol. Chem. 364:1537-1540(1983).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.75 ANGSTROMS).
RC STRAIN=DY;
RX MEDLINE=99041548; PubMed=9826175;
RA Eschenburg S., Genov N., Peters K., Fittkau S., Stoeva S., Wilson K.S., Bezel C.;
RT "Crystal structure of subtilisin DY, a random mutant of subtilisin Carlsberg.";
RL Eur. J. Biochem. 257:309-318(1998).
CC -1- FUNCTION: Subtilisin is an extracellular alkaline serine protease, it catalyzes the hydrolysis of proteins and peptide amides.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity for peptide bonds, and a preference for a large uncharged residue in p1. Hydrolyzes peptide amides.
CC -1- COFACTOR: Binds 2 calcium ions per subunit.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: Secretion of subtilisin is associated with onset of sporulation, and many mutations which block sporulation at early stages affect expression levels of subtilisin. However, subtilisin is not necessary for normal sporulation.
CC -1- SIMILARITY: Belongs to peptidase family S8.
DR PDB; 1BH6; 18-NOV-98.
```

DR MEROPS; S08.037; -;  
 DR InterPro; IPR000209; Peptidase\_S8.  
 DR Pfam; PF00082; Peptidase\_S8; 1.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR PROSITE; PS00136; SUBTILASE ASP; 1.  
 DR PROSITE; PS00137; SUBTILASE HIS; 1.  
 DR PROSITE; PS00138; SUBTILASE\_SRR; 1.  
 KM Hydrolyase; Sporulation; Serine protease; Metal-binding;  
 FT ACT SITE 32 32 CHARGE RELAY SYSTEM.  
 FT ACT SITE 63 63 CHARGE RELAY SYSTEM.  
 FT ACT SITE 220 220 CHARGE RELAY SYSTEM.  
 FT METAL 2 2 CALCIUM 1.  
 FT METAL 41 41 CALCIUM 1.  
 FT METAL 74 74 CALCIUM 1 (VIA CARBONYL OXYGEN).  
 FT METAL 76 76 CALCIUM 1 (VIA CARBONYL OXYGEN).  
 FT METAL 80 80 CALCIUM 1 (VIA CARBONYL OXYGEN).  
 FT METAL 168 168 CALCIUM 2 (VIA CARBONYL OXYGEN).  
 FT METAL 170 170 CALCIUM 2 (VIA CARBONYL OXYGEN).  
 FT METAL 173 173 CALCIUM 2 (VIA CARBONYL OXYGEN).  
 SQ SEQUENCE 274 AA; 27435 MW; 0154696E22F46533 CRC64;

Query Match 4.8%; Score 79.5; DB 1; Length 274;  
 Best Local Similarity 22.3%; Pred. No. 18;  
 Matches 70; Conservative 49; Mismatches 112; Indels 83; Gaps 18;

QY 22 GLRLPGKAKQLQKVEVFPVLDGDCVSIINMGDSVAMTINFLTSLKTPG 81  
 DB 7 GLPLKADKQV-----AQYKGNVAVKGI-----IDGLASHT-----DLKAVG 46  
 QY 82 GSSVGM-----LDLGGSGSTQ-----IAPLPRVEGTLOASPPGYLTALRMFNRTYKLSY 132  
 DB 47 GASFPVSGSYNDGNGHGVAGTVALDNTTGVAVAPNLSLAIKVLNLS----- 98  
 QY 133 LGLGMSARLALIGVEGQPADGKELVSPCL-SSEFGKEHEKAVTRYSGQAAASLH 191  
 DB 99 -GSGTYS--AIVSGIE-WATONGLDVINMSLGGPS-----GSTALKQVDAKAVASGI 146  
 QY 192 ELCAA---RVSEVLQNRVHREVGVDVFVAFSYYYDLAAGVLDAKGG---SLVVD 245  
 DB 147 VVVAANGSGSSGSGNTI-----GYPAYDVIVAGVANDSKNNAASVSGAE 194  
 QY 246 FPIAAKYCRITLFTPOSSPSCMDLTV-----SLLOEGFPRKYLKTRKD 296  
 DB 195 LEVMAFGV-SVYSTVP-SNTYSLNGTSMASPHVAGAAALILSK--YPLTASQVRRLS 250  
 QY 297 NVETSMALGAFHY 310  
 DB 251 STAIN--LGDSPFY 262

RESULT 7  
 YG73\_VIBPA STANDARD; PRT; 287 AA.  
 ID YG73\_VIBPA  
 AC 087LD9;  
 DT 10-OCT-2003 (rel. 42, Created)  
 DT 10-OCT-2003 (rel. 42, Last sequence update)  
 DT 10-OCT-2003 (rel. 42, Last annotation update)  
 DE Hypothetical UPF0042 protein VP2673.  
 GN VP2673.  
 OS *Vibrio parahaemolyticus*.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 CC Vibrionaceae; *Vibrio*.  
 NCBI\_TaxID=670;  
 RN NCBI\_TaxID=670;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RIMD 2210633 / Serotype O3:K6;  
 RX MEDLINE=22508454; PubMed=12620739;  
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,  
 RA Iijima Y., Nishimura M., Nakano M., Yamashita A., Kubota Y., Kimura S.,  
 RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;  
 RT "Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism  
 distinct from that of *V. cholerae*.";

RL Lancel 361.743-749 (2003).  
 CC -1- SIMILARITY: Belongs to the UPF0042 family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AP005082; BAC60936.1; -;  
 DR HAMAP; MF 00636; -; 1.  
 DR InterPro; IPR005337; UPF0042.  
 DR Pfam; PF03668; ATP\_bind2; 1.  
 KM Hypothetical protein; ATP-binding; Complete proteome.  
 FT NP\_BIND 8 15 ATP (POTENTIAL).  
 SQ SEQUENCE 287 AA; 32292 MW; 47B9001DA8550E9A CRC64;

Query Match 4.7%; Score 78; DB 1; Length 287;  
 Best Local Similarity 23.4%; Pred. No. 25;  
 Matches 44; Conservative 32; Mismatches 64; Indels 48; Gaps 10;

QY 116 TALRMENRTYKLSYSLGLGMSARLALIGVEGQPADGKELVSPCLSPFKGEWEHA 175  
 DB 91 TLKRYSETRRIHP--LSLGGQSLSL-----DQAIERREKELTPLKA-----HA 132  
 QY 176 EVTRYSGQKAASLHELCAARVSEVLQNRVH-RTEVGVHDFVAFSYYYDLAAGVLID 234  
 DB 133 DLTIASSGQ---SHEL-----SETVRKVRGRDRKGLVMFESGFRKGLPS-----D 178  
 QY 235 AEKGSGLVWDPEPIAAKYCRITLFTPOSSPSCMDLTVSLLOEGFPRKYLKTRK 294  
 DB 179 A-----DYFVDVRFPLNP-HWEPALRPLTGDAPIAALEQ-----HOSVLKXQ 223  
 QY 295 IDNVETSW 302  
 DB 224 IESFIETW 231

RESULT 8  
 ENGC\_PORGI STANDARD; PRT; 315 AA.  
 ID ENGC\_PORGI  
 AC P59946;  
 DT 15-MAR-2004 (rel. 43, Created)  
 DT 15-MAR-2004 (rel. 43, Last sequence update)  
 DT 15-MAR-2004 (rel. 43, Last annotation update)  
 DE Probable GTPase *engC* (EC 3.6.1.-).  
 GN ENGC OR PG1900.  
 OS *Porphyromonas gingivalis* (Bacteroides gingivalis).  
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;  
 CC Porphyromonadaceae; *Porphyromonas*.  
 NCBI\_TaxID=837;  
 RN NCBI\_TaxID=837;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=W83;  
 RX MEDLINE=22829867; PubMed=12949112;  
 RA Nelson K.E., Fleischmann R.D., Deboy R.T., Paulsen I.T., Fouts D.E.,  
 RA Eisen J.A., Daugherty S.C., Dodson R.J., Durkin A.S., Gwinn M.,  
 RA Haft D.H., Kolonay J.F., Nelson W.C., Mason T., Taiton L., Gray J.,  
 RA Granger D., Tettelin H., Dong H., Galvin J.L., Duncan M.U.,  
 RA Dewhirst F.E., Fraser C.M.;  
 RT "Complete genome sequence of the oral pathogenic bacterium  
 RT *Porphyromonas gingivalis* strain W83.";  
 RL J. Bacteriol. 185:5591-5601(2003).  
 CC -1- FUNCTION: Unusual circulatory permuted GTPase that catalyzes rapid  
 CC hydrolysis of GTP with a slow catalytic turnover (By similarity).  
 CC -1- SIMILARITY: Monomer (Probable).  
 CC -1- SIMILARITY: Contains 1 *engC* GTPase domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; A201718; RAQ6684.1; -  
DR TIGR; PG1900; -; 1;  
DR HAMAP; MF 01820; -; 1;  
DR PROSITE; PS00936; ENCG\_GTPASE; 1;  
KW Hydrolase; GTP-binding; Complete proteome.  
FT DOMAIN 88 241 ENCG\_GTPASE.  
FT NP\_BIND 128 131 GTP (PROBABLE).  
FT NP\_BIND 182 189 GTP (PROBABLE).  
FT NP\_BIND 236 240 GTP (PROBABLE).  
FT SITE 267 280 KNUCKLE-LIKE CYSTEINE CLUSTER.  
FT SEQUENCE 315 AA; 34551 MW; FB98020A040413BE CRC64;

Query Match 4.6%; Score 77; DB 1; Length 315;  
Best Local Similarity 20.1%; Pred. No. 35;  
Matches 69; Conservative 41; Mismatches 101; Indels 132; Gaps 16;

QY 13 TPLVLKATAGRLLPGEKAQLQKVEFKASPFVGVGDVCSIMNGTDEGVSAWI 68

Db 21 TELYCAKGNLR-----KGIRSNPTVVGDR-VEIVPASQDGPVAKIKIH 66

QY 69 -----TINFILTSKTPGGSSVGMGLDGGSTQIAFLPRVEGTLOASPPG 113

Db 67 PRENYIIRASNLKSHILGANL-----DAVLVCTINPTVTTFIDRLATAEAY 119

QY 114 YLTALRMFNRTYKLYS-----YLGIG-----LMSAR 141

Db 120 RVFVILVFN---KIDCYTOEDRLQLDRLSAVTAIGYPCCHVSATGEGPLDKSLDGK 176

QY 142 LATLGGVEGPAKDGKELVSPCLSPFGKEWEHAEVTVRVSGKAAASLHELCAARVSEV 201

Db 177 LTLLAGSGV---GKSLINALIP-----HA-----DLRTGAISQA 209

QY 202 LQNRVHRTVEKHVDVFYAFSYVDYLAAGVLIDAE--KG-GSLVGVDPFI-----AA 250

Db 210 HHTGMHTTFSQIMDF-----PDLSPGALIDTPGKFGFTLMEGEVSHYFPEIFAA 263

QY 251 KYVCR-----TLETQPSPPSCMDLTVVSLLOE 280

Db 264 SKGCRFGNCTHTEPGCAVLEALRRGEIAESRYISVLSILEDE 306

# RESULT 9

BCHC RHOSH STANDARD; PRT; 318 AA.  
AC Q02430; Q9RFC3;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 2-desacetyl-2-hydroxyethyl bacteriochlorophyllide A dehydrogenase (EC 1.1.1.1).  
GN BCHC.  
OS Rhodobacter sphaeroides (Rhodospirillum rubrum sphaeroides).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;  
OC Rhodobacteraceae; Rhodobacter.  
OX NCBI\_TaxID=1063;  
RN [1]  
SEQUENCE FROM N.A. / 2.4.1 / NCIB 8253 / DSM 158;  
PC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;  
RX MEDLINE=93173096; PubMed=8437569;  
RA McGlynn P., Hunter C.N.;  
RT "Genetic analysis of the bchC and bchA genes of Rhodobacter sphaeroides";  
RL Mol. Gen. Genet. 236:227-234 (1993).  
RN [2]  
SEQUENCE FROM N.A. / 2.4.1 / NCIB 8253 / DSM 158;  
RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;  
RX MEDLINE=20115911; PubMed=10648776;  
RA Choudhary M., Kaplan S.;

RT "DNA sequence analysis of the photosynthesis region of Rhodobacter sphaeroides 2.4.1.";  
RL Nucleic Acids Res. 28:862-867 (2000).  
CC -!- FUNCTION: THIS PROTEIN CATALYZES THE PENULTIMATE STEP IN BACTERIOCHLOROPHYLL A BIOSYNTHESIS.

CC -!- PATHWAY: Light-independent bacteriochlorophyll biosynthesis.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AJ010302; CAB38746.1; -  
DR EMBL; AF195122; AAF24296.1; -  
DR PIR; S30914; S30914.  
DR PIR; T50752; T50752.  
DR InterPro; IPR002085; Adh\_zn\_family.  
DR InterPro; IPR005903; BchB.  
DR Pfam; PF00107; Adh\_zinc\_N; 1.  
DR ProDom; PD040557; BchB\_1.  
DR TIGRPFAMs; TIGR01202; bchC; 1.  
KW Photosynthesis; Bacteriochlorophyll biosynthesis; Oxidoreductase.  
FT CONFLICT 278 278 A -> R (IN REF. 1).  
FT CONFLICT 301 302 GL -> AF (IN REF. 1).  
SQ SEQUENCE 318 AA; 33912 MW; C9388CAB94D0E585 CRC64;

Query Match 4.6%; Score 77; DB 1; Length 318;  
Best Local Similarity 21.6%; Pred. No. 35;  
Matches 70; Conservative 40; Mismatches 94; Indels 120; Gaps 19;

QY 22 GLRLPGEKAQLQKVEFKASPFVGVGDVCSIMNGTDEGVSAWITIN 71

Db 58 GYPLVPGVEA-----AGEVVEAAPDTGFRGDRVFPVGSNCFA---PTDAG-----PIR 103

QY 72 FLTGS-----LKTGGSSVGMGLDGGSTQIAFLPRVEGTLOASPPGYLTALRMFNRTYKL 127

Db 104 GLFAATKRLVTPAHRAV-----RIDPALEA--GALLALAAATAR-HAL 144

QY 128 YSYSYL-----GLGLMSARLAILGGVEGPAKDGKELVSPCLSPFGKEWEHAEV 178

Db 145 AGLNHLVPLDLVGHGTLGRLLARLTIAAG--GEP-----FWV-----WE----- 181

QY 179 YRVSGQYAAASLHELCAARVSEVLOQNRVHRTVEKHVDVFYAFSYVDYLAAGVLIDAE--- 235

Db 182 -----TKAERRR-----AGEVEVIDPATDQRRDYRSI-----YDASGDPKLIISLV 224

QY 236 --EKGSLVGVDF---EIAAKYVCRTL-----ETQPSPPSCMDLTVVSLLOEF 281

Db 225 RLAKGGIVLAGFYTFEPAFTVPFAPFMKEARLRIAAEWQPE-----DMVATRALIESG 277

QY 282 GFPRSKVLKTRKIDNVETSWALG 305

Db 278 ALSLANLITTRPASEAAEAYATG 301

## RESULT 10

CLN8 HUMAN STANDARD; PRT; 286 AA.  
ID CLN8\_HUMAN  
AC Q9UBY8; Q96195;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE CLN8 protein.  
GN CLN8.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
SEQUENCE FROM N.A., VARIANT EPMR GLY-24, AND VARIANT ALA-155.

RA MEDLINE=99438402; PubMed=10508524;  
RA Ranta S., Zhang Y., Ross B., Lonka L., Takkenen E., Messer A.,  
RA Sharp J., Wheeler K., Kuusmi K., Mole S., Liu W., Soares M.B.,  
RA Bonaldi M.F., Hirasaniemi A., de la Chapelle A., Gilliam T.C.,  
RA Lehesjoki A.-E.;  
RT "The neuronal ceroid lipofuscinoses in human EPVR and mnd mutant mice  
RT are associated with mutations in CLN8.";  
RT Nat. Genet. 23:233-236(1999).  
RL [2]  
RN SEQUENCE FROM N.A.  
RP TISSUE=Uterus;  
RX MEDLINE=2338257; PubMed=12477932;  
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RX Krausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RX Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RX Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,  
RX Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RX Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,  
RX Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
RX Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mallat S.U.,  
RX Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RX Richards S., Wooley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RX Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RX Fahy J., Hailton E., Kettelman K., Madan A., Rodriguez S., Sanchez A.,  
RX Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RX Rodriguez R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Botterfield Y.S.N., Krzywinski M.I., Skalska U., Smolins D.E.,  
RA Schermer A., Schein J.E., Jones S.J.M., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RL [3]  
RN SUBCELLULAR LOCATION, AND MUTAGENESIS OF 283-LYS-LYS-284.  
RX MEDLINE=20320699; PubMed=10861296;  
RX Lonka L., Kyttälä A., Ranta S., Jalanko A., Lehesjoki A.-E.;  
RT "The neuronal ceroid lipofuscinosis CLN8 membrane protein is a  
RT resident of the endoplasmic reticulum.";  
RT Hum. Mol. Genet. 9:1691-1697(2000).  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic  
CC reticulum and ER-Golgi intermediate compartment (ERGIC).  
CC -1- P.M.: Does not seem to be N-glycosylated.  
CC -1- DISEASE: Defects in CLN8 are a cause of progressive epilepsy with  
CC mental retardation (EPVR) [MIM:600433]; also known as neuronal  
CC ceroid lipofuscinosis type 8 and Northern epilepsy. EPVR is an  
CC autosomal recessive disorder characterized by normal early  
CC development, onset of generalized seizures between 5 and 10 years,  
CC and subsequent progressive mental retardation. Biochemically, the  
CC disease is characterized by the intracellular accumulation of  
CC hydrophobic material, mainly ATP synthase subunit C.  
CC -1- SIMILARITY: Contains 1 TDC (TRAM/LAGL/CLN8) domain.  
CC -1- DATABASE: NMB=NCL CLN8;  
CC NOTE=Neural Ceroid Lipofuscinoses mutation db;  
CC WWW="http://www.ucl.ac.uk/hci/CLN8.html".  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/  
CC or send an email to license@ebi.ac.uk).

DR GO: GO:0005793; C:ER-Golgi intermediate compartment; TAS.  
DR GO: GO:0016021; C:integral to membrane; TAS.  
DR GO: GO:0007399; P:neurogenesis; TAS.  
DR InterPro: IPR00634; TDC.  
DR SMART: SM00724; TDC; 1.  
DR PROSITE: PS00922; TDC; 1.  
KW Transmembrane; Endoplasmic reticulum; Neuronal ceroid lipofuscinosis;  
KW Disease mutation; Polymorphism; Epilepsy.  
FT TRANSMEM 21 41  
FT TRANSMEM 62 84  
FT TRANSMEM 103 123  
FT TRANSMEM 131 151  
FT TRANSMEM 226 246  
FT DOMAIN 62 262  
FT SITE 263 286  
FT VARIANT 24 24  
FT VARIANT 155 155  
FT VARIANT 41  
FT VARIANT 155 155  
FT MUTAGEN 283 284  
FT CONFLICT 225 225  
SQ SEQUENCE 286 AA; 32842 MW; 0BAABBCA51ED1DC CRC64;  
Query Match 4.6%; Score 75.5; DB 1; Length 286;  
Best Local Similarity 24.5%; Pred. No. 42;  
Matches 61; Conservative 28; Mismatches 87; Indels 73; Gaps 14;  
QY 81 GSSVGMGLDGGSTQIAFLPRVEGTLQASPGYLTALRMF-----NRTYK-LYS 129  
DB 7 GGTSSISIDLYAASGI-----KSTLMA--GVFLIGVAVVGHQSSLSNATYRSLVA 58  
QY 130 YSYGLGMSAPRLALGVSGQAPKDGKELVPSLSFGKGEHAEVTVRSQKAAAS 189  
DB 59 REKYFWMDLAATRAVF--GVQSTAGLWMLGDPTV-----HAD--KARQQWMCW 104  
QY 190 LH-----ELCAARYSEVLQVRHTEBEVKVDFAFSYDYDLAAGVGLDAEKGSLV- 242  
DB 105 FHITATGPFCEENVAVHLSLIRFT-----EDLPVYVHHLRAFLGL-----GCLVN 152  
QY 243 --VGDFEIAKVCYCTLETQPSDFSCMDLTVYSLIQEFGFPSPSKYKLTIRKIDNVT 300  
DB 153 LQVG-----HYIANTTLLIEMSTPFC-----VSMMLKAGWSESLFWKLNQ----- 194  
QY 301 SWALGAIFH 309  
DB 195 -WLMTHMFH 202  
RESULT 11  
Y712\_SNY3  
ID Y712\_SNY3 STANDARD; PRT; 322 AA.  
AC Q55979;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DE 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical protein slr0712.  
GN slr0712  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
OX NCBI\_Taxid=1148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96127529; PubMed=8590279;  
RX Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,  
RA Sugita M., Tabata S.; the genome of the unicellular cyanobacterium  
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb  
RT region from map positions 64% to 92% of the genome.";  
RL DNA Res. 2:153-166(1995).  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -1- SIMILARITY: TO E.COLI YBHN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.F., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A.L., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RA laboratory strains.";
RL J. Bacteriol. 184:5479-5490 (2002).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z70283; CAA94259.1; -.
DR EMBL; AE007072; AAK46561.1; -.
DR FIR; D70787; D70787.
DR TIGR; MT2276; -.
DR Tuberculinist; Rv2219; -.
KW Hypothetical protein; Transmembrane; Signal; Complete proteome.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 250 HYPOTHETICAL PROTEIN Rv2219/MT2276.
FT TRANSMEM 51 71 POTENTIAL.
FT TRANSMEM 73 93 POTENTIAL.
FT SEQUENCE 250 AA; 26863 MW; ADC43314243095C CRC64;
SQ
Query Match 4.5%; Score 75; DB 1; Length 250;
Best Local Similarity 22.1%; Pred. No. 39;
Matches 46; Conservative 22; Mismatches 86; Indels 54; Gaps 7;
QY 72 FLTGSLKTPGGSSVGMLDLGGGSGTQIAFLPRVEGTLQASPPGYLTALRPNRTYKLSYS 131
DB 49 YMIGAFLLIVGASVGVGVWAGGFTMTFTMIFL--GVLL---GALVAFVIFGR----- 94
QY 132 YLGLGMSRLAILGVGVGPQAKGKELVSPCLSPFKGEWE-----HAEVTYRVS 182
DB 95 -----RAQRTYTKAEQGTGAANAAL-----DNLGKRWRTPGVAATGNLDVAHRVI 141
QY 183 GKAKAASLHELCAARYSEVLGNRVHRTTEVKHVDFFYAFSYYYDLAAGVGLIDAEKGSLV 242
DB 142 GRGCVTFVGEGSNARKVKPILAQEKKTAR-----LVGDVPIYDIIVGN--- 184
QY 243 VGDFFEIAKYVCRITLTETQPQSPFSCMD 270
DB 185 -GDGEVPLAKLRLRHLTPANITVKQMD 211
RESULT 13
PPNK AQUAE STANDARD; PRT; 274 AA.
ID ID PPNK AQUAE
AC AC O67055;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable inorganic polyphosphate/ATP-NAD kinase (EC 2.7.1.23)
DE (Poly(P)/ATP NAD kinase).
GN PPNK OR AQ.909.
OS Aquifex aeolicus.
OC Bacteria; Aquificae;
CC Aquificales; Aquificaceae; Aquifex.
CX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VP5;
RA MEDLINE=38196666; PubMed=9537320;
RX Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.J., Keller M., AuJay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358 (1998).

```

CC -1- FUNCTION: Catalyzes the phosphorylation of NAD to NADP. Utilizes  
CC ATP and other nucleoside triphosphates as well as inorganic  
CC polyphosphate as a source of phosphorus (By similarity).  
CC -1- CATALYTIC ACTIVITY: ATP + NAD(+) = ADP + NADP(+).  
CC -1- COFACTOR: Divalent metal ions (By similarity).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -1- SIMILARITY: Belongs to the NAD kinase family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AE000713; AAC07026.1; -.  
CC PIR; D70378; D70378.  
CC HAMAP; MF\_00361; -. 1.  
CC InterPro; IPR002504; ATP\_NADK.  
CC Pfam; PF01513; NAD\_kinase; 1.  
CC Transfaser; Kinase; NADP; NADP; Complete proteome.  
CC SEQUENCE 274 AA; 31327 MW; 34510F2B30EB2C6 CRC64;  
SQ  
Query Match 4.5%; Score 74.5; DB 1; Length 274;  
Best Local Similarity 24.3%; Pred. No. 48;  
Matches 42; Conservative 20; Mismatches 64; Indels 47; Gaps 8;  
QY 11 KATPVLKATA-----GRLPGKAKLQK--VKEVFKASPTVDD 52  
DB 97 KVLPLVERAKLQERLMIDVLRNRRLRYIGYLANDAVLSKSIATITKVFINEE 156  
QY 53 CVSINMGDEGVSAWITTFILGSLKTPGSGSVGMIDLG-----GSTOIAFLPREVETL 107  
DB 157 VLEVG---DGV-----ILSTPGSTAYVLSAGPPIVPEGNLFLVPCPTL 202  
QY 108 QASPPGYLTALMNFRTYKLYSYLGIGLSARLALIGVGGQPAKKGKELV 160  
DB 203 SNRP--LVLPSPKFEVFKVSENV-----MEAFLL-LDGGEGFLLKKGDEVI 245  
RESULT 14  
PYRA\_METTF STANDARD; PRT; 227 AA.  
ID PYRA\_METTF  
AC P29577;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 01-APR-1993 (Rel. 25, Last annotation update)  
DE Hypothetical 24.7 kDa protein (ORF5A).  
OS Methanobacterium thermoformicicum.  
OG Plasmid pFV1.  
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
OC Methanobacteriaceae; Methanothermobacter.  
OX NCBI\_TaxID=145262;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DSM 3848 / THF;  
RX MEDLINE=93126090; PubMed=1336177;  
RA Noelling J., van Breden F.J.M., Eggen R.I.L., de Vos W.M.;  
RT "Molecular organization of related Archaeal plasmids encoding different  
RT restriction-modification systems in Methanobacterium  
RT thermoformicicum";  
RT Nucleic Acids Res. 20:6501-6507 (1992).  
CC -1- SIMILARITY: TO ORF5 IN PFZ1.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X68366; CA448429.1; -.  
DR PIR; S30305; S26440. DUF11.  
DR InterPro; IPR001434; DUF11.  
DR Pfam; PF01345; DUF11.1.  
DR TIGRFAMs; TIGR01451; B\_ant\_repeat; 1.  
KM Hypothetical protein; Plasmid.  
SQ SEQUENCE 227 AA; 24681 MW; 2F30042519E849E CRC64;  
Query Match 4.4%; Score 73.5; DB 1; Length 227;  
Best Local Similarity 23.1%; Pred. No. 46;  
Matches 45; Conservative 29; Mismatches 64; Indels 57; Gaps 11;  
QY 64 VSAW-ITINELTSLKTPG-----SSVGMIDLGSGSTOIAFLPREVGTQASPPGYL 115  
DB 54 VKAMNLDVGNATARVQLPEGLVQDYVMGQYDLEGTWEIGDIPAYE-----ERSL 106  
QY 116 TALMFRRTYKLYSYLGIGLSARLALIGVGGQPADGKELVPCLSPTKGEWEA 175  
DB 107 TFIQLNRT-----GSVTVANVTADDDNSANNMEL-----TFK-VFGIS 147  
QY 176 ETTYRVSQKAAALSHELCAAYSEVLQNR-VARTEEVKVDPEYA-----FSY---YY 224  
DB 148 DLEVNTGKNETARIGD--TVRTVTLKRRGHIDANNIKGNFLSGGLVQNFSDAGIF 205  
QY 225 D-----IAAG 229  
DB 206 DDITREWIFETLLAAG 220  
RESULT 15  
PCNA\_PYRAB STANDARD; PRT; 249 AA.  
ID PCNA\_PYRAB  
AC Q9UYX8;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE DNA polymerase sliding clamp (Proliferating cell nuclear antigen  
DE homolog) (PCNA).  
GN PCN OR PYRAB13790 OR PAB1465.  
OS Pyrococcus abyssi.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Pyrococcus.  
OX NCBI\_TaxID=29292;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GBS / Orsay;  
RX MEDLINE=22511545; PubMed=12622808;  
RA Cohen G.N., Barbe V., Flament D., Galperin M., Hellig R., Lecompte O.,  
RA Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,  
RA Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;  
RT "An integrated analysis of the genome of the hyperthermophilic  
RT archaeon Pyrococcus abyssi";  
RT Mol. Microbiol. 47:1495-1512 (2003).  
CC -1- FUNCTION: Sliding clamp subunit. Responsible for tethering the  
CC catalytic subunit of DNA polymerase to DNA during high-speed  
CC replication (By similarity).  
CC -1- SUBUNIT: Homotrimer (By similarity).  
CC -1- SIMILARITY: Belongs to the PCNA family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AJ248287; CAB50284.1; -.  
CC PIR; G75048; G75048.  
CC HAMAP; MF\_00317; -. 1.  
CC InterPro; IPR000730; Pr\_cel\_nuc\_antig.  
CC Pfam; PF00705; PCNA; 1.

```

DR Pfam: PF02747; PCNA_C; 1.
DR PRINTS; PR00339; PCNACYCLIN.
DR ProDom; PD002673; Pr_cel_nuc_antig; 1.
DR TIGRFAMs; TIGR00590; pcna; 1.
DR PROSITE; PS01251; PCNA_1; 1.
DR PROSITE; PS00293; PCNA_2; FALSE NEG.
KW DNA-binding; DNA replication; Complete proteome.
SQ SEQUENCE 249 AA; 28035 MW; B7E21A135B3F3BC8 CRC64;

Query Match      4.4%; Score 73.5; DB 1; Length 249;
Best Local Similarity 22.0%; Pred. No. 52;
Matches 49; Conservative 31; Mismatches 100; Indels 43; Gaps 10;

QY 33 KLLQKVEYFKASPLVGGDDCVSINNGTDEGVSAITINF---LGSLSKTPGGSVG--- 86
Db 15 QLIETASRLIDEAAFAKVEEGIS-NRAMPSPRVVLIDNLPASIFSKYEVDGEETIGVNM 73
QY 87 -----MLDLGGGSTQIAFLPRVEGTLOASPPCYLTALRMFNRTYK--LYSYSYLGGL-- 137
Db 74 DHLKVKLKGKAKETILLRKGEENFLEISLQGTAT-----RTFKPLPIDVEEIEVDLPE 127
QY 138 --MSARLAILGGVEGQPAKDGKELYSPCLSPSKGMEHAEVTVYVSGKAAASLHELCA 195
Db 128 LPFTAKVILGDVTKBAVKD-----ASLVSDSMKFIAKENFTMAREG-----ETQE 174
QY 196 ARVSEVLQNR---VHRTVEVKHVDYFAFSYYDYDLAAGVGLID 234
Db 175 VEVKLTLEDEGLLDIEVQETKSA--YGISYLSDMVKGLGKAD 215

```

Search completed: June 8, 2004, 10:09:08  
Job time : 21 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: June 8, 2004, 10:06:02 ; Search time 45 Seconds

(without alignments)  
2257,708 Million cell updates/sec

Title: US-09-905-589A-2\_COPY\_135\_456

Perfect score: 1659  
Sequence: 1 AKODIPDFMKATPLVLKAT.....ALGAIHYIDSLNFKSPAS 322

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 651375

Minimum DB seq length: 0

Maximum DB seq length: 322

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_25:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_virus:\*

16: sp\_bacteriaph:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	643.5	38.8	278	11 O8CH23	O8CH23 mus musculu
2	377	22.7	224	11 O8CDB6	O8CDB6 mus musculu
3	374.5	22.6	271	5 O8IPZ6	O8IPZ6 drosophila
4	115.5	7.0	300	11 O8K012	O8K012 mus musculu
5	113	6.8	322	16 O8G316	O8G316 bifidobacte
6	94	5.3	280	16 O69808	O69808 streptomyce
7	88	5.7	246	16 O821C9	O821C9 chlamydomon
8	86	5.2	280	16 O821R0	O821R0 streptomyce
9	85.5	5.2	244	16 O9X810	O9X810 streptomyce
10	85	5.1	322	16 O8ESU0	O8ESU0 oceanobacti
11	84	5.1	293	16 O9WZ05	O9WZ05 thermocoga
12	84	5.1	321	16 O7W077	O7W077 bordetella
13	84	5.1	321	16 O7W1W1	O7W1W1 bordetella
14	83.5	5.0	224	16 O82CJ2	O82CJ2 streptomyce
15	83.5	5.0	246	11 O8BVP6	O8BVP6 mus musculu
16	83.5	5.0	280	16 O8PAR8	O8PAR8 xanthomonas

17	83	5.0	304	2 O7MWV3	O7MWV3 alcaligenes
18	82	4.9	212	2 O52738	O52738 rhodococcus
19	82	4.9	258	10 O84K39	O84K39 oryza sativ
20	82	4.9	267	10 O7XV22	O7XV22 oryza sativ
21	82	4.9	297	16 O92U54	O92U54 rhizobium m
22	82	4.9	322	5 O9XK39	O9XK39 caenorhabdi
23	81.5	4.9	267	16 O8U677	O8U677 agrobacteri
24	81.5	4.9	276	5 O43942	O43942 leishmania
25	81.5	4.9	312	5 O9BJU5	O9BJU5 leishmania
26	81.5	4.9	321	16 O7VS81	O7VS81 bordetella
27	81	4.9	183	13 O7ZUJ5	O7ZUJ5 brachydanio
28	81	4.9	223	16 O97JCI	O97JCI clostridium
29	81	4.9	279	17 O8TLR4	O8TLR4 methanobact
30	80.5	4.9	237	3 O8XIC6	O8XIC6 talaromyces
31	80.5	4.9	312	5 O9GUB0	O9GUB0 leishmania
32	80	4.8	277	16 O8INQ9	O8INQ9 bacillus an
33	80	4.8	280	16 O821O7	O821O7 streptomyce
34	79.5	4.8	264	16 O8F7N9	O8F7N9 leptospira
35	79	4.8	322	5 O22524	O22524 caenorhabdi
36	78.5	4.7	245	17 O9YAR7	O9YAR7 aeropyrum p
37	78.5	4.7	312	5 O43944	O43944 leishmania
38	78.5	4.7	313	2 O84HP4	O84HP4 amycolatops
39	78	4.7	286	16 O8RBL2	O8RBL2 thermobacter
40	78	4.7	308	16 O8CX80	O8CX80 oceanobacti
41	77.5	4.7	176	2 O49142	O49142 methyllobact
42	77.5	4.7	278	2 O9EXX2	O9EXX2 enterococcu
43	77.5	4.7	284	13 O90348	O90348 coturnix co
44	77.5	4.7	314	16 O24922	O24922 helicobacte
45	77.5	4.7	319	17 O8TY61	O8TY61 methanopyru

## ALIGNMENTS

RESULT 1	ID	Q8CH23	PRELIMINARY;	PRT;	278 AA.
AC	O8CH23				
DT	01-MAR-2003 (T-EMBLrel.. 23, Created)				
DT	01-JUN-2003 (T-EMBLrel.. 23, Last sequence update)				
DE	Similar to ectonucleoside triphosphate diphosphohydrolase 6.				
GN	ENTPD6.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
OX	NCBI_TaxId=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=Mammary gland;				
RA	Strausberg R.;				
RU	Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; BC038126; AAH38126.1; -				
DR	MGI; MGI:1202295; Enpd6.				
DR	GO; GO:0016787; F:hydrolase activity; IEA.				
DR	InterPro; IPR000407; GDAI_CD39_NTPase.				
DR	Pfam; PF01150; GDAI_CD39; 1.				
KW	Hydrolase.				
SQ	SEQUENCE 278 AA; 30130 MW; 75A92DD1AC76297F CRC64;				
Query Match	38.8%;	Score 643.5;	DB 11;	Length 278;	
Best Local Similarity	88.1%;	Pred. No. 7.1e-50;			
Matches 126;	Conservative	7;	Mismatches 5;	Indels 5;	Gaps 1;
OY	1 AKODIPDFMKATPLVLKATAGRLLPGEKAGTLOKYEVRKASPTLVGDCCVIMNGT 60				
DB	135 AKGHIPDFMKATPLVLKATAGRLLPGEKAGTLOKYEVRKASPTLVGDCCVIMNGT 194				
OY	61 DEVSAMITINFLTGLKTPGGSSVGMIDLGGGSGTQIAFLPVEGTLQASPPGYVLTALRM 120				
DB	195 DEVSAMITINFLTGLKTPGGSSVGMIDLGGGSGTQITFLPVEGTLQASPPGYVLTALRM 254				
OY	121 FRTYKLVSYSLGLGMSARLA 143				



```

Db      255 FNRTKLYSYRW-----VCSRLA 272

RESULT 2
ID Q8CDB6 PRELIMINARY; PRT; 224 AA.
AC Q8CDB6;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Ectonucleoside triphosphate diphosphohydrolase 6.
EN ENTPD6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK030789; BAC27138.1; -.
DR MGD; MGI:1202295; Entp6d.
DR InterPro; IPR000407; GDAL CD39_NTPase.
DR Pfam; PF01150; GDAL CD39_1.
SQ SEQUENCE 224 AA; 24538 MW; 305DCC373B5B4A8E CRC64;

Query Match 22.7%; Score 377; DB 11; Length 224;
Best Local Similarity 96.0%; Pred. No. 7.3e-26;
Matches 72; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AKQIIPDFWKATPLVLKATAGLRLLPGSKAQLKQKEVFKASPFVLGDDCVSINMGT 60
135 AKQHIIPDFWKATPLVLKATAGLRLLPGSKAQLKQKEVFKASPFVLGDDCVSINMGT 194
Db      61 DGVSAWITINFLTG 75
QY 135 DGVSAWITINFLTG 209
Db      195 DGVSAWITINFLTG 209

RESULT 3
Q8IP26 PRELIMINARY; PRT; 271 AA.
AC Q8IP26;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DE C3059-FC.
GN NTPASE OR C3059.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20196006; PubMed=10731132;
RX Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Gabor G.L.,
RA Abiril J.F., Agbayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

```

```

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Farriera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinart K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.N., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang Q., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC Celnikier S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Jacul J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D., Venter J.C.,
RA Williams S.N., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochmk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celnikier S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celnikier S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AEO03591; AAN10397.1; -.
DR FlyBase; FBgn0024947; NTPase.
DR InterPro; IPR000407; GDAL CD39_NTPase.
DR Pfam; PF01150; GDAL CD39_1.
SQ SEQUENCE 271 AA; 30025 MW; BF01FF322360A363 CRC64;

Query Match 22.6%; Score 374.5; DB 5; Length 271;
Best Local Similarity 36.8%; Pred. No. 1.6e-25;
Matches 103; Conservative 41; Mismatches 101; Indels 35; Gaps 12;

```

```

QY 52 DCVSMNGTDEGVSMITINFLTGL- KTPGSSVGMIDLGGSTQIAFLPRVETGLQAS 110
DB 2 DAVEIMDGTDEGIFSMFTVNFGLRLSKT---NQAPADLDGGSTQVTFSP---TDPDQ 54
QY 111 PGY-----LTLRMFRRTKYLSYSYGLGLMSARLALIGVGEPADGKELVSPC 163
DB 55 VPVYKRVHEVVTSSKIN---VTHSYLGLMAAHVNF--THGY-KKEDIVLESVC 107
QY 164 LSPSKGE-WEHAETVTVSGQKAAASLHE-----LCAARY-SEVLQNRVHRTVEYKH 214
DB 108 VPIIANRTWYGVNYQYKVSCKENKSSAEOPIVDPAELVKSVMPLVKPFTLKQ 167
QY 215 VDFVAFSYVYDLAAVGLIDAEKGSILVGDPELTAKVCTLEHQPSSSPSCMDLTYV 274
DB 168 HAVAAFSYFESALISGLVDPLAGGETVEAVRKAQELCAIPDE---QPFMCFDLTFI 224
QY 275 SLILQE-FGPRSKVLKLRKIDNVETSMALGALFHYIDS 313
DB 225 STLREGFLNDGKKIKLYKKIDGHEISMALGCAVNVLT 264

```

## RESULT 4

```

Q8K0L2 PRELIMINARY; PRT; 300 AA.
AC Q8K0L2;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DB RIKEN CDNA 2010320H07 gene.
GN 2010320H07R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC031143; AAH31143.1;
DR MGD; MGI:1919340; 2010320H07R1K.
DR InterPro; IPR000407; GDAI_CD39_NTPase.
DR Pfam; PF01150; GDAI_CD39; 1.
SQ SEQUENCE 300 AA; 33645 MW; C33E372AFB266B1C CRC64;

```

Query Match 7.0%; Score 115.5; DB 11; Length 300;  
 Best Local Similarity 20.7%; Pred. No. 0.056; Mismatches 84; Indels 103; Gaps 12;  
 Matches 59; Conservative 39;

```

QY 85 VGMIDLGGSTQIAFLPR---VEGLQASPPGYLTALMERNRYKLYSYSLGLMSAR 141
DB 3 VGADLDLGASTQISFVPGPIIDOSTQVT-----FLYQANYSVYTHSYLFCG---R 51
QY 142 LAIIGVGOPAKXG-KELVSPCLSPSKGEWEHAETVTVSGQKAAASLHE-----CAA 196
DB 52 DQILNRILAKLAQDRLSQVAPVRHPCVH-----SGYQALLPLSLSLDSEFC-- 97
QY 197 RVSEVLQNRVHRTVEYKH----- 215
DB 98 -----HTTDSINHTQNLVEGTGDPGNCVVALRLFNFSCKQKQCAFNGLYOP 148
QY 216 -----DEVAFSYVYDLAAVGLIDAEKGSILVGDPELTAKVCTLEHQPSSSPSCMDLTYV 267
DB 149 PVHGFVAFSNFTYTHFLNLTSRQSLNV---NDTYWKFQCGKPMKLVESYSPGQERWL 204
QY 268 ---CMDLTYVSLILQEFGRPSKV---LKLTRKIDNVETSMALG 305
DB 205 RDYCASGLYLIVLLE-GYKFSSEETWPIQFOKQAGNVDICWTLG 248

```

RESULT 5  
 Q8G3L6 PRELIMINARY; PRT; 322 AA.  
 ID Q8G3L6

```

AC Q8G3L6;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DB Widely conserved hypothetical protein.
GN B11742.
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCC 2705;
RX MEDLINE=22294977; PubMed=12381787;
RA Schell M.A., Karimiantzou M., Snel B., Vilanova D., Berger B.,
RA Pessi G., Zwielen M.-C., Desiere F., Bork P., Delley M.,
RA Pridmore R.D., Arigoni F.;
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract."
RI Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427 (2002).
DR EMBL; AB014608; JAM25526.1;
DR InterPro; IPR003695; Ppx-GppA.
DR Pfam; PF02541; Ppx-GppA; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 322 AA; 35198 MW; 10BB35585F54BC87 CRC64;

```

Query Match 6.8%; Score 113; DB 16; Length 322;  
 Best Local Similarity 24.7%; Pred. No. 0.11; Indels 40; Gaps 7;  
 Matches 47; Conservative 29; Mismatches 74;

```

QY 10 WKATPVLKATAGLRLLPEKQKLYQKVEYFKASPFVAGDVCVSMNGTDEGVSAWIT 69
DB 78 YEITQLLWATSLKLRAP--NKNVYHKIE-----LIQG-VTVLSGTDEA-----R 122
QY 70 INFVLSLKTGGSSVG---MIDLGGSTQIAFLPRVETGLQASPPGYLTALMERNRYK 126
DB 123 LTFE--NARRWYGMWDGRLLVLDIGGSLVAMGSDDEPTVALSAPA----- 167
QY 127 LVYSYSLGLMSARLALIGVGOPAKGKELVSPCLSPSKGEWEHAETVTVSGQK 186
DB 168 -----GAGRTTELPBGMAATPDELDVRKQVKILBPMKVVPQSGRPHAVGTSKT 220
QY 187 AASLHELCA 196
DB 221 FSLRLCGA 230

```

## RESULT 6

```

ID 069808 PRELIMINARY; PRT; 280 AA.
AC 069808;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DB Hypothetical protein SC06414.
GN SC06414 OR SC1A6.03.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Parkhill J., Bentley S.D., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);

```

```

RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.; and a detailed genetic and physical map for
RT "A set of ordered cosmids and a detailed genetic and physical map for
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RN Mol. Microbiol. 21:77-96(1996).
RN
RN SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=2196410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wierozorek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL93127; CRA18901.1; -.
DR PIR; T28684; T28684.
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.
DR GO; GO:0006807; P:nitrogen metabolism; IEA.
DR InterPro; IPR003010; NtIse/CNhydase.
DR Pfam; PF00785; CN_hydrolase; 1.
DR PROSITE; PS50263; CN_HYDROLASE; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 280 AA; 31804 MW; CE8E483E6E40B167 CRC64;

Query Match 5.7%; Score 94; DB 16; Length 280;
Best Local Similarity 31.1%; Pred. No. 4.5;
Matches 37; Conservative 13; Mismatches 53; Indels 16; Gaps 5;

QY 147 GVGQPAKD---GKELVSPCLSPFKGEW-----EHAETVTVSGQKAAAS-----LHEL 193
DB 136 GNSGWPVFDVAVKGVYICVDHFFEGWALGLEAEIVFNPSPATSGLSGYLVQLEGP 195

QY 194 CAARVSEVLQNRVHR--TEEVGHVDYAFSYVDLAAG-VGLIDAEKGLSVGVGFEIA 249
DB 196 AAANVANEYFGAINRVGVEELGNDVFGTGYFVDPPEARFVGEVASKETELVVRDLDA 254

RESULT 7
Q821C9 PRELIMINARY; PRT; 246 AA.
AC Q821C9;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Virulence protein pgp6-D.
GN CCAA00001.
OS Chlamydomophila caviae.
OG Plasmid pCpGP1.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83557;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=GPIC;
RX MEDLINE=22569155; PubMed=12682364;
RA Read T.D., Myers G.S.A., Brurham R.C., Nelson W.C., Paulsen I.T.,
RA Heidelberg J., Holtzapfel E., Khouri H., Federova N.B., Carty H.A.,
RA Unayen L.A., Haft D.H., Peterson J., Beanan M.J., White O.,
RA Salzberg S.L., Hsia R.-C., McClarty G., Rank R.G., Bavali P.M.,
RA Fraser C.M.;
RT "Genome sequence of Chlamydomophila caviae (Chlamydia psittaci GPIC):
RT examining the role of niche-specific genes in the evolution of the
RT Chlamydiaceae.";
RL Nucleic Acids Res. 31:2134-2147(2003).
DR EMBL; AE015926; AAP05752.1; -.
DR TIGR; CCAA00001; -.

```

```

DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR InterPro; IPR005350; UPF0137.
DR Pfam; PF03677; UPF0137; 1.
DR Plasmid; Complete proteome.
SQ SEQUENCE 246 AA; 28045 MW; 1B4C3F239179BF43 CRC64;

Query Match 5.3%; Score 88; DB 16; Length 246;
Best Local Similarity 21.1%; Pred. No. 13;
Matches 43; Conservative 40; Mismatches 77; Indels 44; Gaps 10;

QY 21 AGRLDGEKAKLQKVEVFKASFLVGD-----DCVSINMGTDGVSANITINFLT 74
DB 65 ASIKLLTGG-----IKSIQHVLLIGEKYKREILRAVNSPDTTFSAINLVFHT 116
QY 75 GSKLTPGSGVGMLDLGGSGTQIAFLP--RVEGTLOASPPGYLTALRMENRTYKLYSYS 132
DB 117 ---KSSAYNALGYEL-----FISLPDKDTKSLFQSIP--YKTAYLLASRK----- 157
QY 133 LGGLMGAAILGGVEGQPAKQKELVSPCLSPFKGEWEHAETVTVSGQKAAASLHE 192
DB 158 ---GSVXDKVKVLGKIEGNTAAIDILNRL--PSLRASQTERSINFPEDKXKSEKIME 213
QY 193 L---CAA-RVSEVLQNRVHRTE 211
DB 214 ILKTVCSGLSELYNKLLQQLFE 237

RESULT 8
Q821R0 PRELIMINARY; PRT; 280 AA.
AC Q821R0;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative hydrolase.
GN SAV1950.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL; AP005028; BAC69661.1; -.
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.
DR GO; GO:0006807; P:nitrogen metabolism; IEA.
DR InterPro; IPR003010; NtIse/CNhydase.
DR Pfam; PF00795; CN_hydrolase; 1.
DR PROSITE; PS50263; CN_HYDROLASE; 1.
DR Complete proteome.
SQ SEQUENCE 280 AA; 31553 MW; B25558BD9A269A43 CRC64;

Query Match 5.2%; Score 86; DB 16; Length 280;
Best Local Similarity 29.4%; Pred. No. 24;
Matches 35; Conservative 14; Mismatches 54; Indels 16; Gaps 5;

```

QY 147 GVEGQPAKD---GKEIVSPCLSPSKGEME-----HAEVTVRVSGCKAAS-----LHEL 193  
DB 136 GNSGMPFEDTAVGIVGTYICTDRHFPESGMRALGLAGBIVFNPNATRGISGYIMQLECP 195  
QY 194 CAATSEVTLQNRVHR--TEEVKHYDFAFSTYYDLAAG-VGLIDAEKGSGLVGDPEIA 249  
DB 196 AAATVANEVFGAINRVGVEELGDNDFGTSYFVDPEAFQVEGVAASDKETELVVRDLDLA 254

## RESULT 9

Q9X8L0 PRELIMINARY; PRT; 244 AA.  
ID Q9X8L0; Q9X8L0;  
AC Q9X8L0;  
DT 01-NOV-1999 (Tremblrel. 12, Created)  
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)  
DE Possible membrane protein.  
GN SC03431 OR SCE9.38.  
OS Streptomyces coelicolor.  
OC Bacteria; Actinobacteriales; Actinomycetales;  
OC Streptomycinae; Streptomycetaceae; Streptomyces.  
RX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Oliver K., Harris D.;  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Redenbach M., Kiese R.H.M., Denapalte D., Eichner A., Cullum J.,  
RX MEDLINE=97000351; PubMed=8843436;  
RA Kinsht H., Hopwood D.A.;  
RT "A set of ordered cosmid and a detailed genetic and physical map for  
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
RL Mol. Microbiol. 21:77-96(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / M145;  
RX MEDLINE=21996410; PubMed=12000953;  
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kiese R.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.H., Kiese R., Larke L., Murphy L., Oliver K., O'Neill S.,  
RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete Streptomyces  
coelicolor A3(2)." ;  
RL Nature 417:141-147(2002).  
RN EMBL, AL938116; CAB42784.1; --  
DR PIR, T36357; T36357.  
DR GO; GO:0005215; F:transporter activity; IEA.  
DR GO; GO:0006810; F:transport; IEA.  
DR InterPro; IPR000566; Lipocin cytochrome P450.  
DR PROSITE; PS00213; LIPOCALIN\_1.  
KW Complete proteome.  
SQ SEQUENCE 244 AA; 23480 MW; CADDBEL140F940EA CRC64;  
Query Match 5.2%; Score 85.5; DB 16; Length 244;  
Best Local Similarity 23.3%; Pred. No. 22;  
Matches 53; Conservative 23; Mismatches 92; Indels 59; Gaps 9;  
QY 67 WTIITFLNGSLKTPGGSGVGMGLDGGSTQIAFLPRVEGTLOASPPG-----LTALRMF 121  
DB 20 WVTAGNTASTENTASBESTS---GAMRTGYA-----AGTASATDPGTDRSGGVAFFRS 70

QY 122 NRTYKLYSYSLGILGMSARLIL-----GGVEGQPAKDGEIVSPCLSPS----- 167  
DB 71 RRTFRSRRVVAALSLGLVGA-LALITACTGGSDGGSGSGGSDGSSAPGASASTDGG 129  
QY 168 -----FKGEW-----EHAETVTVRVSGCKA--ASLHELCAATSEVTLQNRVHR 208  
DB 130 DAGGSPSAGAGELGSGMLATTGQAVALMTVGDGKALFATGCTVCSGTTEETSGTFRIR 189  
QY 209 -----TEEVKHYDFAFSTYYDLAAGVGLIDAEKGSGLVYG 244  
DB 190 LKCADGSADPRATGKAVGATSLTVAMGALGKETVTRSEGSLLPBG 236

## RESULT 10

Q9ESJ0 PRELIMINARY; PRT; 322 AA.  
ID Q9ESJ0; Q9ESJ0;  
AC Q9ESJ0;  
DT 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Thiamine monophosphate kinase (EC 2.7.4.16).  
GN OB0644.  
OS Oceanobacillus iheyensis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.  
OX NCBI\_TaxID=182710;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HTEB31 / DSM 14371 / JCM 11309;  
RX MEDLINE=22280767; PubMed=12235376;  
RA Takami H., Takaki Y., Uchiyama I.;  
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya  
Ridge and its unexpected adaptive capabilities to extreme  
environments." ;  
RL Nucleic Acids Res. 30:3927-3935(2002).  
DB EMBL: AP004595; BAC12600.1; --  
DR GO; GO:0009330; P:thiamin-phosphate kinase activity; IEA.  
DR GO; GO:0016740; P:transferase activity; IEA.  
DR GO; GO:0009228; P:thiamin biosynthesis; IEA.  
DR InterPro; IPR000728; AIR synth.  
DR InterPro; IPR006283; ThL.  
DR Pfam; PF00586; AIRS; 1.  
DR Pfam; PF02769; AIRS\_C; 1.  
DR TIGRfam; TIGR01379; thl; 1.  
KW Kinase, transferase, Complete proteome.  
SQ SEQUENCE 322 AA; 35872 MW; 6E711ADCD0DE6B763 CRC64;  
Query Match 5.1%; Score 85; DB 16; Length 322;  
Best Local Similarity 21.0%; Pred. No. 36;  
Matches 69; Conservative 44; Mismatches 122; Indels 94; Gaps 14;

QY 35 LQKVEVFRASPFL--VDDCVSINMGTDGVSAMTITNFTGSLKTPGGSGVGMGLDG 91  
DB 7 ISIKPTTYRQPSLKGIDDDDAVFRSSKDIYAVDTFEDVHFRSTWEPYHIGYRALG 66  
QY 92 GGSTQIAFLPRVEGTLOASPPGYLALRM-FRITYKLYSYSLGILGMSAR-LALIGS- 147  
DB 67 ANISDLA-----AMGASPSFLYSIVPKMTQALSQITSGHDDLASKXHMILIGD 119  
QY 148 -VEG-----QPADGK-ELVSPCLSPSKG----- 170  
DB 120 TVSGKQLTISITVIGYITNKARVROHAKGDIVFVGTGLGSLAGFHILTSNENRXY 179  
QY 171 -----EMEAETVTVRVSGCKAASLHELCAATSEVTLQNRVHRTEVKVH----- 215  
DB 180 DQDFYTHRRKPEPRVAFQALBSLDRVSLNDVSGIANEASLIVASVSVTLAENDIP 239  
QY 216 ---DFAFS---YYDLAAGVGLIDAEKGSGLVGDPEIAKRYCRTLLETPOGSSPFGC 268  
DB 240 VAPSFYQFTYEQGYQKLSGG--DFELLGTAKGDVAVVNEAAKKT-NTG----- 287  
QY 269 MDITVYSLLOEFGRP-----RSKRYLK 290

```
Db 288 --LTWIGSVVEEKHPVYIEDNAMRKVLK 314
RESULT 11
O9WZQ5
ID O9WZQ5 PRELIMINARY; PRT; 293 AA.
AC O9WZQ5;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Malonyl CoA-acyl carrier protein transacylase.
GN TM0798.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
Raft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
McDonald L., Utterback T.R., Malek J.A., Linner K.D., Garrett M.M.,
Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RA "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-325(1999).
DR EMBL; AB001747; AAD35880.1; -.
DR PIR; G72334; G72334.
DR HSSP; P25715; 1MLA.
DR TIGR; TM0798; -.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0004314; F:[acyl-carrier protein] S-malonyltransferase. .; IEA.
DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001227; Ac.trans.
DR InterPro; IPR004410; Fabd.
DR Pfam; PF00698; Acyl.transf. 1.
DR TIGRFAMs; TIGR00128; fabd; 1.
DR Complete proteome.
KW SEQUENCE 293 AA; 32966 MW; F98A4F033A88194 CRC64;
Query Match 5.1%; Score 84; DB 16; Length 293;
Best Local Similarity 24.2%; Pred. No. 39;
Matches 59; Conservative 32; Mismatches 67; Indels 86; Gaps 13;
Qy 25 LIPGEKAO-----KLQKVEFKASPFVLGDDCVSINMGTDGV-----SAWI 68
Db 5 VFPQGSQYSGWAKDFSVYESSKEIFERSKVLGFDITEIMMGDEETKLTENAQPSYI 64
Qy 69 TINFILGSLTPGSSVGMGLDGGGSTQIAFLPRVEGTLOASPPGYLTALRMFRNTKLY 128
Db 65 T-----SYIAFLEL-----EKRGILPDV---VAGHSIGYTAALAVAG-VYDFE 103
Qy 129 SVSYL-----GLGMSARLAILGGVGPQAKGKELVSPCLSPFKGEW-- 172
Db 104 TGLYLVKRGGEYMSKALEPKGTMAA-----VIGLNITETIEEVN-----SIEGVIA 151
Qy 173 ---EHAEVTVRVSQ-----OKLAASHELCAARVSEVLQNRVHRT-----EVRK 213
Db 152 NYNSHQVV--ISGLKESVEKAMEILKEGARVVELVMSVSPFHTPFLEYAREKKEEVE 209
Qy 214 HVDF 217
Db 210 KVDF 213
RESULT 12
Q7WQ7
ID Q7WQ7 PRELIMINARY; PRT; 321 AA.
AC Q7WQ7;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
```

```
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Putative exported protein.
GN BB0238.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RB50 / ATCC BAA-588;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Saunders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrett B.G., Maskell D.J.;
RA "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640437; CAE30736.1; -.
KW Complete proteome.
KW SEQUENCE 321 AA; 33666 MW; DEF695A0B678F43E CRC64;
Query Match 5.1%; Score 84; DB 16; Length 321;
Best Local Similarity 22.8%; Pred. No. 45;
Matches 69; Conservative 37; Mismatches 136; Indels 60; Gaps 14;
Qy 17 LKATAGLRLLPGKAKOL-LQKVEFKASPFVLGDDCVSINMGTDGVSAMTINFLTG 75
Db 8 LLALALTPAAGAAQNWFSQPLRWIV---PYPAG-----GGTD-----VVARTVAG 51
Qy 76 SLKTPGSSVGMGLDGGGSTQIAFLPRVEGTLOASPPGYL-----TALRMFRNTKLYSY 130
Db 52 SLEKPLGQTTIVENRPGAAITIG---ATAIAQADPAGYVVGVTADSGTLAFNPSL----- 102
Qy 131 SYLGLGMSARLAILGGVGPQAKGKELVSPCLSPFKGEWEHAEVTVRVSGQKAAASL 190
Db 103 -YAKLSYDPAKFTYIGGIARFPLMLAVNNSPKSVEDVLQARKEPAKLTASAGAGSP 161
Qy 191 HELCAARVSEVLQNRVHRTVEVKHVDYAFS-----YYDLAGVGLIDAEK 237
Db 162 HHLAL---ELFKQRAN--VDVLHVPYKGAAPAIQDLGGQVDMMPIDLAAGLPNI---K 212
Qy 238 GGSL-VVGDFEIAAKVCTRLETQPOSSPFCMDLTIVSL--LLQEFGEFRSKVLKLTAK 294
Db 213 AGKRLVLG---TATPERLAVLPDAPTMASQGVADFTAYAWQGLVGPAGMPEAAVKKLGGE 269
Qy 295 ID 296
Db 270 LE 271
RESULT 13
Q7W1W1
ID Q7W1W1 PRELIMINARY; PRT; 321 AA.
AC Q7W1W1;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Putative exported protein.
GN BPP0234.
OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=519;
RN [1]
RP SEQUENCE FROM N.A.
```

RC STRAIN=12822 / ATCC BAA-587;  
RX MEDLINE=22827954; PubMed=12910271;  
RA Parkhill J., Sebastia M., Preston A., Murphy L.D., Thomson N.,  
RA Harris D.E., Holden M.T.G., Churcher C.M., Harris B., Quail M.A.,  
RA Cerdeno-Tarrega A.M., Temple L., James K., Harris B., Quail M.A.,  
RA Achtman M., Atkin R., Baker S., Basham D., Baeson N., Cherevach I.,  
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,  
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holtroyd S., Jagsels K.,  
RA Leathley S., Moulis S., Norbertczak H., O'Neill S., Ormond D., Price C.,  
RA Rabinowitch E., Rutter S., Sanders M., Saunders S., Stevens K.,  
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,  
RA Unwin L., Whitehead S., Barrett B.G., Maskell D.J.,  
RT "Comparative analysis of the genome sequences of *Bordetella pertussis*,  
RT *Bordetella parapertussis* and *Bordetella bronchiseptica*.";  
RL Nat. Genet. 35:32-40(2003).  
DR EMBL: BX640423; CAZ39975.1; -  
KM Complete proteome.  
SQ SEQUENCE 321 AA; 33666 MW; DEF695A0B678F43E CRC64;

Query Match 5.1%; Score 84; DB 16; Length 321;  
Best Local Similarity 22.8%; Pred. No. 45;  
Matches 69; Conservative 37; Mismatches 136; Indels 60; Gaps 14;

QY 17 LKATAGRLPGEKAKL-LQKVEVFKASFLVGDCCVIMNGTDEGVSAITINFLNG 75  
DB 8 LLAALALTFAAGAAQWPSQPLKMTV--PYAG-----GSTD-----VVARTVAG 51  
QY 76 SLKTPGSSVGMLDLGGSTQIAFLPRVEGTLQASPPGYL-----TALRMFNTYKLYSY 130  
DB 52 SLKRLPGTIVENRPGATITIG---ATAIAQDPAGYVVGATDSTLAFNLSL----- 102  
QY 131 SYLGGLMSAFALILGVEGQPAQDQKELVSPCLSPFKSEWHEAVTVTVSGQKAAAL 150  
DB 103 -YAKLSYDPAKFTYIGGARFPLMAVAVNSPYKSVEDVLQAKKEPAKLTASAGAGSP 161  
QY 191 HELCAARVSEVLQNRVTRTEVKHVDYFAS-----YYVDLAGGLIDAEK 237  
DB 162 HHLAL---ELFKORAN--VDLHVPYKGAAPALQDLGGQVDMFIDLAAGLPNT---K 212  
QY 238 GGSF-VVGDFFIAKAYVCRLTETPOSSPSCMDLTVSL--LLOEGFPRSKVLKTRK 294  
DB 213 AGKRVIG---TAPERLAVLPDAPVMAEQGVADFTAYVAGLVGPAQMEBAVKKLGGE 269  
QY 295 ID 296  
DB 270 LE 271

RESULT 14  
Q82CJ2 PRELIMINARY; PRT; 224 AA.

AC Q82CJ2  
DT 01-JUN-2003 (TREMblrel. 24, Created)  
DT 01-JUN-2003 (TREMblrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN SAV5353.  
OS Streptomyces avermitilis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycinae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=33903;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RX MEDLINE=21477403; PubMed=11572948;  
RA Omura S., Ikeda H., Ishikawa J., Hamamoto A., Takahashi C., T.,  
RA Shinoe M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,  
RA Kikuchi H., Sakai Y., Hattori M.,  
RT "Genome sequence of an industrial microorganism *Streptomyces*  
RT *avermitilis*: deducing the ability of producing secondary  
RT metabolites.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RX MEDLINE=22608306; PubMed=12692562;  
RA Ikeda H., Ishikawa J., Hamamoto A., Shinoe M., Kikuchi H., Shiba T.,  
RA Sakai Y., Hattori M., Omura S.,  
RT "Complete genome sequence and comparative analysis of the industrial  
RT microorganism *Streptomyces avermitilis*.";  
RL Nat. Biotechnol. 21:526-531(2003).  
DR EMBL: AP005042; BAC73065.1; -  
DR GO: GO:0016020; C:membrane; IEA.  
DR GO: GO:0009103; P:lipopolysaccharide biosynthesis; IEA.  
DR InterPro: IPR003856; LPS\_M2\_MPA.  
DR Pfam: PF02706; wzz\_1.  
KM Hypothetical protein; Complete proteome.  
SQ SEQUENCE 224 AA; 22584 MW; B7EF5BF696917CC CRC64;

Query Match 5.0%; Score 83.5; DB 16; Length 224;  
Best Local Similarity 28.1%; Pred. No. 29;  
Matches 47; Conservative 15; Mismatches 80; Indels 25; Gaps 6;

QY 90 LGGSTQIAFLPRVEGTLQASPPGY-LTALRMFNTYKLYSYLGL-----GLMSARLAI 144  
DB 33 LAAAGTLAAGAGAVGL--TPPAYSATAYVIAVPTATSDQALGLQAVGRVATGLAV 90  
QY 145 LGGVE---GQPKDQKELVSPCLSPFKSEWHEAVTVTVSGQKRAALHLCARVSEV 201  
DB 91 LGDAQVMAGVPKTLRENVQATSPD-----APMVAVATATSRADLATDMANVARS 142  
QY 202 LQNRVTRTEVKHVDYFASYYTDLA-----AGVGLIDAEKGSYL 241  
DB 143 LIRHNDIKMATHVELLQFSRAVRKAPSASPALTGLVGASAGLL 189

RESULT 15  
Q8BVP6 PRELIMINARY; PRT; 246 AA.

AC Q8BVP6  
DT 01-MAR-2003 (TREMblrel. 23, Created)  
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)  
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)  
DE Hypothetical serine proteases (Hypothetical protein).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; Tissue=Testis;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The PANTOM Consortium,  
RA the PAXTON Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC Tissue=Testis;  
RA Strausberg R.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AK077071; BAC36592.1; -  
DR EMBL: BC049744; AAH49744.1; -  
DR GO: GO:0004295; P:tryptase activity; IEA.  
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro: IPR001254; Peptidase S1.  
DR PROSITE: PS00134; TRYPSIN\_HIS\_1.  
KM Hypothetical protein.  
SQ SEQUENCE 246 AA; 26996 MW; 9D5FA13F10CC279D CRC64;

Query Match 5.0%; Score 83.5; DB 11; Length 246;  
Best Local Similarity 21.8%; Pred. No. 33;  
Matches 53; Conservative 36; Mismatches 89; Indels 65; Gaps 11;

QY 10 WKATPLVLTAKTGLPGEKAKLQKVEVFKASFLVGDCCVIMNGTDEGVSAITINFLNG 69

Db	6	WRSLOLLYLEA-ISLPCTEALLCYEATASAFRA-----VSLHNW-----KWLL	49
QY	70	I-----NFLTGLKT-----PGSSVGMLDLGGSTQIAFLPRVEGTLOASPPGYLTALRM	120
Db	50	LRSMVCNOREGCEETVFIETGTSKVLSPKGCSSAFSPYQI--SYLVSPPGVSI-----	104
QY	121	FNRTYKLYSYSLGLGL-----MSARLAILGGVEGOPAKDGKE-----	158
Db	105	---SYSRVCRSYLCNNLTNLEFFVRLKASQPMSTLPSAKSCPCSCVKGHDQCLPSFVYTE	161
QY	159	-----LVSPCLSPSPKGEWEHAEVTVRVG-----OKAAASLHELCAARVSEVLQNRVHRT	209
Db	162	NCPFAASSCYSTLKFQAGNLNTTFLINGCARDSHKLLADFOHIGSIRTEVI-NVLDKS	220
QY	210	EVV	212
Db	221	EAV	223

Search completed: June 8, 2004, 10:10:06  
Job time : 48 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: June 8, 2004, 10:02:16 ; Search time 58 Seconds

(without alignments)  
1568.626 Million cell updates/secTitle: US-09-905-589a-2 COPY\_135\_456  
Perfect score: 1659  
Sequence: 1 AKQIDPFDFMKATPLVLKAT.....ALGALFHYIDSLNRQKSPAS 322Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1278565

Minimum DB seq length: 0  
Maximum DB seq length: 322Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1980s:\*  
3: geneseqp2000s:\*  
4: geneseqp2000s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	596	35.9	141	AAU30879	AAU30879 Novel hum
2	596	35.9	144	AAU30881	AAU30881 Novel hum
3	189.5	11.4	243	AAQ22051	AAQ22051 Arabidops
4	186	11.2	139	AAU30880	AAU30880 Novel hum
5	184	11.1	217	AAQ22052	AAQ22052 Arabidops
6	151.5	9.1	269	AAE01419	AAE01419 Human gen
7	151.5	9.1	288	AAE01356	AAE01356 Human gen
8	151.5	9.1	288	ABG64123	ABG64123 Human alb
9	113	6.8	322	ABP66314	ABP66314 Bifidobac
10	104	6.3	46	AAU44306	AAU44306 Peptide #
11	104	6.3	46	ABR33253	ABR33253 Peptide #
12	104	6.3	46	AAU26776	AAU26776 Peptide #
13	104	6.3	46	ABR28079	ABR28079 Human pep
14	104	6.3	46	ABR18715	ABR18715 Protein #
15	104	6.3	46	AAU66436	AAU66436 Human Don
16	104	6.3	46	AAU54045	AAU54045 Human bra
17	104	6.3	46	ABG48102	ABG48102 Human liv
18	104	6.3	46	AAU02035	AAU02035 Peptide #
19	104	6.3	46	ABG36085	ABG36085 Human pep
20	86	5.2	279	ABR59533	ABR59533 Drosophi
21	86	5.1	281	ABR67409	ABR67409 Drosophi
22	84.5	5.1	274	AAU14887	AAU14887 Subtilisi
23	84.5	5.1	274	AAU14866	AAU14866 Subtilisi
24	84.5	5.1	274	AAU14874	AAU14874 Subtilisi
25	83.5	5.0	274	AAU14880	AAU14880 Subtilisi

26	83	5.0	212	2	AAU14161	AAU14161 R. rhodoc
27	83	5.0	212	2	AAU02292	AAU02292 Beta ch
28	83	5.0	293	2	AAU28961	AAU28961 Serrat
29	82.5	5.0	192	5	ABP63032	ABP63032 Human pol
30	82.5	5.0	198	5	ABP41815	ABP41815 Human ova
31	82.5	5.0	274	2	AAU14886	AAU14886 Subtilisi
32	81.5	4.9	243	4	ABG01294	ABG01294 Novel hum
33	81.5	4.9	274	2	AAU14868	AAU14868 Subtilisi
34	81.5	4.9	283	6	ABP57689	ABP57689 Saccharop
35	81.5	4.9	299	7	ADBE64638	ADBE64638 Human pro
36	81	4.9	212	2	AAU13898	AAU13898 Nitrite h
37	81	4.9	293	2	AAU07057	AAU07057 Serrat
38	80.5	4.9	274	1	ABP80849	ABP80849 Sequence
39	80.5	4.9	274	2	AAU14872	AAU14872 Subtilisi
40	80.5	4.9	274	2	AAU14878	AAU14878 Subtilisi
41	80.5	4.9	274	2	AAU14854	AAU14854 Subtilisi
42	80.5	4.9	274	2	AAU14863	AAU14863 Subtilisi
43	80.5	4.9	274	2	AAU14871	AAU14871 Subtilisi
44	80.5	4.9	274	2	AAU14851	AAU14851 Subtilisi
45	80.5	4.9	274	2	AAU14856	AAU14856 Subtilisi

## ALIGNMENTS

RESULT 1  
AAU30879  
ID AAU30879 standard; protein, 141 AA.  
XX  
XX AAU30879;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DB Novel human secreted protein #1370.  
XX  
XX Human; vaccination; gene therapy; nutritional supplement;  
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;  
KV immune suppression; immune stimulation; anti-inflammatory; leukaemia.  
XX  
OS Homo sapiens.  
XX  
XX WO200179449-A2.  
XX  
PD 25-OCT-2001.  
XX  
XX 16-APR-2001; 2001WO-US008656.  
XX  
XX 18-APR-2000; 2000US-00552929.  
XX  
XX 26-JAN-2001; 2001US-00770160.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT;  
XX  
XX WPI; 2001-611725/70.  
XX  
XX Nucleic acids encoding a range of human polypeptides, useful in genetic  
PT vaccination, testing and therapy.  
XX  
XX Claim 20; Page 365; 765pp; English.  
XX  
XX The invention relates to novel human secreted polypeptides. The  
XX polypeptides and antibodies to the polypeptides are useful for  
XX determining the presence of or predisposition to a disease associated  
XX with altered levels of polypeptide. The polypeptides are also useful for  
XX identifying agents (agonists and antagonists) that bind to them. Cells  
XX expressing the proteins are useful for identifying a therapeutic agent  
XX for use in treatment of a pathology related to aberrant expression or  
XX physiological interactions of the polypeptide. Vectors comprising the  
XX nucleic acids encoding the polypeptides and cells genetically engineered  
XX to express them are also useful for producing the proteins. The proteins  
XX are useful in genetic vaccination, testing and therapy, and can be used  
XX as nutritional supplements. They may be used to increase stem cell



CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon  
CC and/or nerve tissue growth or regeneration; immune suppression and/or  
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.  
CC AAU29510-AAU3304 represent the amino acid sequences of novel human  
CC secreted proteins of the invention  
XX  
XX  
SQ Sequence 141 AA;

Query Match 35.9%; Score 596; DB 4; Length 141;  
Best Local Similarity 77.1%; Pred. No. 1.7e-52;  
Matches 121; Conservative 2; Mismatches 14; Indels 20; Gaps 1;  
  
Qy 163 CLSPFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTTEVKHVDYFAFSY 222  
Db 5 CLSPAF-----ILLECAARVKEGLPNRVHRTTEVKHVDYFAFSY 44  
  
Qy 223 YDIAAGVGLIDAEKGSLLVWGDPEFAAKVCTLTQPSQSPFSCMDLTYSLLIQEKG 282  
Db 45 YDLAGAGPIDAEKGSLLVWGDPEFAATKVCVCTLTQPSQSPFSCMDLTYSLLIQEKG 104  
  
Qy 283 FPRSKVLKLRKIDNVETSWALGAIFHYIDSLNRQKS 319  
Db 105 FPRSKVLKLRKIDNVETSWALGAIFHYIDSLNRQKS 141

RESULT 2  
AAU30881  
ID AAU30881 standard; protein; 144 AA.  
XX  
AC AAU30881;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Novel human secreted protein #1372.  
XX  
KW Human; vaccination; gene therapy; nutritional supplement;  
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;  
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.  
XX  
OS Homo sapiens.  
XX  
PN WO200179449-A2.  
XX  
PD 25-OCT-2001.  
XX  
PF 16-APR-2001; 2001WO-US008656.  
XX  
PR 18-APR-2000; 2000US-00552929.  
PR 26-JAN-2001; 2001US-00770160.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT;  
XX  
XX WPI; 2001-611725/70.  
XX  
XX Nucleic acids encoding a range of human polypeptides, useful in genetic  
XX vaccination, testing and therapy.  
XX  
XX Claim 20; Page 365; 765pp; English.  
XX  
XX The invention relates to novel human secreted polypeptides. The  
XX polypeptides and antibodies to the polypeptides are useful for  
XX determining the presence of or predisposition to a disease associated  
XX with altered levels of polypeptide. The polypeptides are also useful for  
XX identifying agents (agonists and antagonists) that bind to them. Cells  
XX expressing the proteins are useful for identifying a therapeutic agent  
XX for use in treatment of a pathology related to aberrant expression or  
XX physiological interactions of the polypeptide. Vectors comprising the  
XX nucleic acids encoding the polypeptides and cells genetically engineered  
XX to express them are also useful for producing the proteins. The proteins  
XX are useful in genetic vaccination, testing and therapy, and can be used  
XX as nutritional supplements. They may be used to increase stem cell

CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon  
CC and/or nerve tissue growth or regeneration; immune suppression and/or  
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.  
CC AAU29510-AAU3304 represent the amino acid sequences of novel human  
CC secreted proteins of the invention  
XX  
XX  
SQ Sequence 144 AA;

Query Match 35.9%; Score 596; DB 4; Length 144;  
Best Local Similarity 76.2%; Pred. No. 1.8e-52;  
Matches 122; Conservative 3; Mismatches 1; Indels 34; Gaps 2;  
  
Qy 163 CLSPFKGEWEHAEVTVRVSGQKAAASLHELCAARVSEVLQNRVHRTTEVKHVDYFAFSY 222  
Db 5 CLAP-----AASLHELCAAKVSEVLHNRVHRTTEVKHVDYFAFSY 44  
  
Qy 223 YDIAAGVGLIDAEKGSLLVWGDPEFAIAKY-----VCRLETQPOSSPFSC 268  
Db 45 YDIAAGVGLIDAEKGSLLVWGDPEFAIAKYGVVTSWVKGRVSSPCVCRLETQPOSSPFSC 104  
  
Qy 269 MDLTYSVLLQEFQFPRSKVLKLRKIDNVETSWALGAIF 308  
Db 105 MDLTYSVLLQEFQFPRSKVLKLRKIDNVETSWALGAIF 144

RESULT 3  
AAG22051  
ID AAG22051 standard; protein; 243 AA.  
XX  
AC AAG22051;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 24833.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; Genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-00301439.  
XX  
PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125789P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 16-APR-1999; 99US-0128714P.  
PR 19-APR-1999; 99US-0129845P.  
PR 21-APR-1999; 99US-0130077P.  
PR 23-APR-1999; 99US-0130449P.  
PR 28-APR-1999; 99US-0130510P.  
PR 28-APR-1999; 99US-0130691P.  
PR 30-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 04-MAY-1999; 99US-0132407P.  
PR 05-MAY-1999; 99US-0132484P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.

PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144684P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.

PR 27-JUL-1999; 99US-0145818P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151068P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151338P.  
PR 01-SEP-1999; 99US-0151330P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155559P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156595P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0158293P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.

```
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161932P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 11.4%; Score 189.5; DB 3; Length 243;
Best Local Similarity 24.7%; Pred. No. 1.5e-10;
Matches 60; Conservative 43; Mismatches 91; Indels 49; Gaps 8;

Qy 103 VEGTLQSPGGLTALRMFNRTYKLYSYGLGLMSARLAILGGVGPAXDGKELVSP 162
Db 17 VEG-----ESYVREMYIKGRKFLVHSYLHYGLAARAIL-----KVEDSNRP 63

Qy 163 CLSPFKEGWEHAEVTVRSVGKAAALHBLCAARVSEVLQNRVHRTTEVKHV----- 215
Db 64 CIATGVAGTYKYGKAFKAAASPGASLDBCRVAINALKVNSLCT---HMKCTFGGV 119

Qy 216 -----DFYAFSYVYDLAAGVGLIDAEKGSVVG-DEIRAKYVCRITLETQPS 264
Db 120 WNGGGGGGQKMFVASFPPDRAAEAGFVDNPQVAEVRPLDPEKAANKACNRMMEGRSK 179

Qy 265 -----PFCWDLTY-VSLLOBFPGFPRSKVLKLRKID-----NVETSWALGAIFHY 310
Db 180 FPRVEEDNLFVCLDLVYQTVLLVDGFLKPSQITILVKVKYGDVAVEAANPLGSAIEA 239

Qy 311 IDS 313
Db 240 VSS 242

RESULT 4
AAU30880
ID AAU30880 standard; protein; 139 AA.
AC AAU30880;
XX
XX
XX 18-DEC-2001 (first entry)
XX
XX Novel human secreted protein #1371.
XX
XX Human; vaccination; gene therapy; nutritional supplement;
XX stem cell proliferation; haematopoiesis; nerve tissue regeneration;
XX immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
XX Homo sapiens.
XX
XX WO200179449-A2.
XX
XX 25-OCT-2001.
XX
XX 16-APR-2001; 2001WO-US008656.
XX
XX 18-APR-2000; 2000US-00552929.
XX
XX 26-JAN-2001; 2001US-00770160.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-611725/70.
XX
XX Nucleic acids encoding a range of human polypeptides, useful in genetic
XX vaccination, testing and therapy.
XX
XX Claim 20; Page 365; 765pp; English.
XX
XX The invention relates to novel human secreted polypeptides. The
XX polypeptides and antibodies to the polypeptides are useful for
XX determining the presence of or predisposition to a disease associated
```

```
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising the
CC nucleic acids encoding the polypeptides and cells genetically engineered
CC to express them are also useful for producing the proteins. The proteins
CC are useful in genetic vaccination, testing and therapy, and can be used
CC as nutritional supplements. They may be used to increase stem cell
CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC AAU29510-AAU33304 represent the amino acid sequences of novel human
CC secreted proteins of the invention
XX
XX Sequence 139 AA;

Query Match 11.2%; Score 186; DB 4; Length 139;
Best Local Similarity 75.0%; Pred. No. 1.4e-10;
Matches 39; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

Qy 1 AKQDIPDFWKATPLVKATAGLLELPGEKAQKLLQKVKESPFELVGDD 52
Db 87 AKQDIPDFWKATPLVKATAGLLELPGEKAQKLLQKVKESPFELVGDD 139

RESULT 5
AAU22052
ID AAU22052 standard; protein; 217 AA.
XX
XX AAG22052;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 24834.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
XX
XX 05-MAR-1999; 99US-0123180P.
XX
XX 09-MAR-1999; 99US-0123588P.
XX
XX 23-MAR-1999; 99US-0125788P.
XX
XX 29-MAR-1999; 99US-0126264P.
XX
XX 01-APR-1999; 99US-0126785P.
XX
XX 06-APR-1999; 99US-0128234P.
XX
XX 08-APR-1999; 99US-0128714P.
XX
XX 16-APR-1999; 99US-0129845P.
XX
XX 19-APR-1999; 99US-0130077P.
XX
XX 21-APR-1999; 99US-0130449P.
XX
XX 23-APR-1999; 99US-0130510P.
XX
XX 28-APR-1999; 99US-0130891P.
XX
XX 30-APR-1999; 99US-0131449P.
XX
XX 30-APR-1999; 99US-0132048P.
XX
XX 30-APR-1999; 99US-0132407P.
XX
XX 04-MAY-1999; 99US-0132484P.
XX
XX 05-MAY-1999; 99US-0132485P.
XX
XX 06-MAY-1999; 99US-0132486P.
XX
XX 07-MAY-1999; 99US-0132487P.
XX
XX 11-MAY-1999; 99US-0132863P.
XX
XX 14-MAY-1999; 99US-0134256P.
XX
XX 14-MAY-1999; 99US-0134218P.
XX
XX 14-MAY-1999; 99US-0134219P.
```

PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137522P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 14-JUN-1999; 99US-0138847P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 17-JUN-1999; 99US-0139453P.  
PR 18-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147202P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148584P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 22-AUG-1999; 99US-0149902P.  
PR 22-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151803P.  
PR 31-AUG-1999; 99US-0151838P.  
PR 01-SEP-1999; 99US-0151830P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155348P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158832P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0158993P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.

PR	25-OCT-1999;	99US-0161405P.	
PR	25-OCT-1999;	99US-0161406P.	
PR	26-OCT-1999;	99US-0161359P.	
PR	26-OCT-1999;	99US-0161360P.	
PR	26-OCT-1999;	99US-0161361P.	
PR	28-OCT-1999;	99US-0161920P.	
PR	28-OCT-1999;	99US-0161992P.	
PR	28-OCT-1999;	99US-0161993P.	
PR	29-OCT-1999;	99US-0162142P.	
Query Match			
Best Local Similarity 11.1%; Score 184; DB 3; Length 217;			
Matches 56; Conservative 40; Mismatches 83; Indels 44; Gaps 7;			
QY	123	RYKLYSYLGLMSARLAILGVGEQPAKDGLVPCLSFSGEHEHAEVTVRS	182
DB	6	RYFLVHSLYHLGLAAAEIL-----KVSDDNNPCIAATGYAGTYKYGKAFKA	57
QY	183	GQKAASLHELCAARVSEVLQNRVHRTVEVKHV-----DFVAFSYVD	225
DB	58	APSGASLDECRVAINALKVNNLSCT---HMKCTFGVWNGGGGQKQKMFVASFDD	113
QY	226	LAAGVGLIDAEKGSVLWG-DPEIAAKVYCTLETQPOSS-----PFSQNDLYV-V	274
DB	114	RAAEAGFVDPNQVAFVRPLDPEKAANKACNNRMEEGSKFPRVEEDNLPYLCILDLYQY	173
QY	275	SLLQBFQFPRSKVLKLRKID---NVETSWALGAIFHYIDS	313
DB	174	TLVDGFGKPSQTITLVKKVYGYAVEAAWPLGSAIEAVSS	216
RESULT 6			
AAE01419			
ID	AAE01419	standard; protein; 269 AA.	
AC	AAE01419;		
XX	18-JUL-2001	(first entry)	
DE	Human gene 5	encoded secreted protein fragment, SEQ ID NO:143.	
XX	Human;	secreted protein; proliferative disorder; cancer; tumour; asthma;	
KW	foetal abnormality; developmental abnormality; haematopoietic disorder;		
KW	immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;		
KW	Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;		
KW	psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;		
KW	inflammation; neurological disorder; Alzheimer's disease; food additive;		
KW	angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;		
KW	pregnancy-related disorder; endocrine disorder; infection; wound healing;		
KW	cell culture; chemotaxis; vulnarity; binding partner identification;		
XX	gene therapy.		
XX	Homo sapiens.		
OS			
XX			
EH	Key	Location/Qualifiers	
FT	Misc-difference 87	/note= "Corresponds to any of the naturally occurring L-amino acids"	
FT	Misc-difference 99	/note= "Corresponds to any of the naturally occurring L-amino acids"	
FT	Misc-difference 230	/note= "Corresponds to any of the naturally occurring L-amino acids"	
FT	Misc-difference 263	/note= "Corresponds to any of the naturally occurring L-amino acids"	
FT	Misc-difference 264	/note= "Corresponds to any of the naturally occurring L-amino acids"	
FT	Misc-difference 264	/note= "Corresponds to any of the naturally occurring L-amino acids"	
XX	WO200134629-A1.		

DE Human gene 5 encoded secreted protein HDPH52, SEQ ID NO:78.  
 XX Human; secreted protein; proliferative disorder; cancer; tumour; asthma;  
 KM foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KM immune system disease; AIDS; autoimmune disease; Rheumatoid arthritis;  
 KM Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;  
 KM psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;  
 KM inflammation; neurological disorder; Alzheimer's disease; food additive;  
 KM angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;  
 KM pregnancy-related disorder; endocrine disorder; infection; wound healing;  
 KM cell culture; chemotaxis; vulnerability; binding partner identification;  
 KM gene therapy.  
 XX Homo sapiens.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /label= Signal\_peptide  
 FT Protein 22..288  
 FT /label= Mature\_human\_secreted\_protein  
 FT Misc-difference 87  
 FT /label= Unknown  
 FT /note= "Encoded by GKT"  
 FT Misc-difference 99  
 FT /label= Unknown  
 FT /note= "Encoded by GKT"  
 FT Misc-difference 230  
 FT /label= Unknown  
 FT /note= "Encoded by SGA"  
 FT Misc-difference 263  
 FT /label= Unknown  
 FT /note= "Encoded by YCT"  
 FT Misc-difference 264  
 FT /label= Unknown  
 FT /note= "Encoded by CYC"  
 FT Misc-difference 270  
 FT /label= Unknown  
 FT /note= "Encoded by GNT"  
 FT Misc-difference 275  
 FT /label= Unknown  
 FT /note= "Encoded by NNA"  
 FT WO200134629-A1.  
 PN 17-MAY-2001.  
 XX  
 XX 08-NOV-2000; 2000WO-US030654.  
 PF 12-NOV-1999; 99US-0164835P.  
 PR 27-JUL-2000; 2000US-0221142P.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA Ruben SM, Komatsoulis GA, Wei P, Fiscella M, Baker KP;  
 PI WPI; 2001-308779/32.  
 DR N-PSDB; AAD05224.  
 XX  
 XX New nucleic acid encoding one of 21 human secreted proteins for  
 PT diagnosing, preventing, treating or ameliorating medical conditions, such  
 PT as autoimmune disease and cancer, and used as a food additive or  
 PT preservative.  
 PS  
 PS Claim 11; Page 436-437; 490pp; English.  
 XX  
 XX AAD05220-AAD05282 represent cDNAs corresponding to 21 human secreted  
 CC protein genes, and ABE01352-ABE01413 represent the proteins they encode.  
 CC AAB01415-AAB01433 represent human secreted protein fragments or variants.  
 CC The secreted proteins and their genes are useful for preventing, treating  
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.  
 CC Pathological conditions can be diagnosed by determining the amount of the  
 CC new protein in a sample or by determining the presence of mutations in  
 CC the new genes. Specific uses are described for each of the 21 genes,

CC based on the tissues in which they are most highly expressed, and include  
 CC developing products for the diagnosis or treatment of proliferative  
 CC disorders; cancer, tumours, foetal and developmental abnormalities;  
 CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,  
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
 CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,  
 CC angiogenic disorders, kidney disorders, gastrointestinal disorders,  
 CC pregnancy-related disorders, endocrine disorders, and infections. The  
 CC proteins can also be used to aid wound healing and epithelial cell  
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs  
 CC before transplantation, for supporting cell culture of primary tissues,  
 CC to regenerate tissues, to identify their cognate ligands or binding  
 CC partners, and in chemotaxis, and can be used as a food additive or  
 CC preservative to modify storage properties. Antibodies specific for a  
 CC protein of the invention can be used in alleviating symptoms associated  
 CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
 CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The  
 CC present sequence represents a human secreted protein of the invention  
 XX  
 XX Sequence 288 AA;  
 SQ  
 Query Match 9.1%; Score 151.5; DB 4; Length 288;  
 Best Local Similarity 35.1%; Pred. No. 1.4e-06;  
 Matches 40; Conservative 14; Mismatches 45; Indels 15; Gaps 2;  
 Qy 1 AKODIPDFPKATPVLVKAAGRLPGERAKTL-OKYAEVFKASPFVGDGCVSIMG 59  
 Db 155 AAHAVPKGKRPVLYICTNAGRLPERKQALADYVDLPEDFLFSQGAVISG 214  
 Qy 60 TDEGSAMITINFL-----TGLTPGSSVGMIDLGSGSTQIAF 99  
 Db 215 KQEGVAMIGINFLVLRFDHEDSDHAETQELAAGRRTVIGIDMGAXXQIAV 268  
 RESULT 8  
 ID ABE64123 standard; protein; 288 AA.  
 XX  
 XX ABE64123;  
 AC  
 XX 27-AUG-2002 (first entry)  
 DT  
 XX  
 XX Human albumin fusion protein #798.  
 DE  
 XX  
 XX Albumin fusion protein; therapeutic protein X; human albumin; HA;  
 KM human serum albumin; HSA; cancer; reproductive disorder;  
 KM digestive disorder; immune disorder; endocrine disorder;  
 KM haematopoietic disorder; neural disorder; connective disorder;  
 KM cytosolic; antifertility; antiinflammatory; antilicer;  
 KM immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;  
 KM neuroprotective; antiParkinsonian; antimicrobial; neuroleptic;  
 KM osteopathic; antiarthritic.  
 XX  
 XX Homo sapiens.  
 OS  
 OS Synthetic.  
 XX  
 XX WO200177137-A1.  
 EN  
 XX  
 XX 18-OCT-2001.  
 PD  
 XX  
 XX 12-APR-2001; 2001WO-US011988.  
 PF  
 XX  
 XX 12-APR-2000; 2000US-0229358P.  
 PR 23-APR-2000; 2000US-019384P.  
 PR 21-DEC-2000; 2000US-0256931P.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX  
 XX Rosen CA, Haseltine WA;  
 PI  
 XX  
 XX WPI; 2002-010886/01.

XX New fusion protein for treating disease e.g. diabetes comprises an  
PT albumin fused to a therapeutic protein.  
XX  
XX Claim 1; Page 1026-1027; 2102pp; English.  
XX  
CC The present invention relates to albumin fusion proteins comprising a  
CC therapeutic protein X and human albumin (HA), also known as human serum  
CC albumin, HSA). The proteins are useful for treating a disease or disorder  
CC that may be modulated by therapeutic protein X. The albumin extends the  
CC shelf-life of protein X, and may increase its biological in vitro/in vivo  
CC activity. The protein is useful for treating and diagnosing disorders  
CC such as cancer, reproductive disorders, digestive disorders (e.g. Crohn's  
CC disease, ulcerative colitis), immune disorders (e.g. acquired  
CC immunodeficiency syndrome, AIDS), endocrine disorders (e.g. diabetes),  
CC haematopoietic disorders, neural disorders (e.g. Alzheimer's,  
CC Parkinson's, Creutzfeldt-Jacob disease, encephalomyelitis, meningitis,  
CC schizophrenia) and connective disorders (e.g. osteoporosis, arthritis).  
CC ABG63326-ABG65518 represent albumin fusion proteins of the invention  
XX  
SQ Sequence 288 AA;  
Query Match 9.1%; Score 151.5; DB 5; Length 288;  
Best Local Similarity 35.1%; Pred. No. 1.4e-06;  
Matches 40; Conservative 14; Mismatches 45; Indels 15; Gaps 2;  
Qy 1 AKQIPDFWKPATPLVLTAGLRLPGEKAQKLL-QKVEVFKASFFLVGDDCVSIMNG 59  
Db 155 AAHPVVKCHETPLTYLTAGMRLPPEKQALAILADLVKDLPLEDFLFSQSQAENVISG 214  
Qy 60 TDEGVSAWITNFI-----TGSLKTPGGSSVGMGLDGGSGSTQIAF 99  
Db 215 KQGVYAWIGINFLVFXRPHDESDAEATQELAGRRVVGILDMGAXXQIAY 268  
RESULT 9  
ABP66314  
ID ABP66314 standard; protein; 322 AA.  
XX  
AC ABP66314;  
XX  
DT 19-NOV-2002 (first entry)  
XX  
DE Bifidobacterium longum NCC2705 ORF amino acid sequence SEQ ID NO:1058.  
XX  
KW Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;  
KW antidiarrheic; antibacterial; inhibitor of Salmonella; detection;  
KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;  
KW rotavirus; food composition; pharmaceutical composition.  
XX  
OS Bifidobacterium longum.  
XX  
PN EP1227152-A1.  
XX  
PD 31-JUL-2002.  
XX  
PF 30-JAN-2001; 2001EP-00102050.  
XX  
PR 30-JAN-2001; 2001EP-00102050.  
XX  
PA (NEST) SOC PROD NESTLE SA.  
XX  
DR WPI; 2002-668397/72.  
XX  
XX Novel polynucleotide comprising Bifidobacterium genome sequence useful as  
PT a probe or primer for detecting and/or identifying Bifidobacterium longum  
PT in a biological sample.  
XX  
XX Claim 3; SEQ ID NO 1058; 80pp; English.  
XX  
CC The present invention describes a polynucleotide (I) comprising a  
CC sequence of a Bifidobacterium genome selected from the nucleotide  
CC sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at

CC least 90% identity or which hybridises with the sequences given in  
CC ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a  
CC fusion protein, comprising a sequence selected from 1097 sequences given  
CC in ABP65258 to ABP6354 ligated in frame to a polynucleotide encoding a  
CC heterologous polypeptide. (I) has antidiarrheic and antibacterial  
CC activities, and can be used as an inhibitor of Salmonella. (II) which is  
CC a probe) is useful for the detection and/or identification of  
CC Bifidobacterium longum in a biological sample. A carrier containing the  
CC lactic acid bacterium Bifidobacterium longum NCC2705 (NCIM 1-2618) can be  
CC used for preventing and/or treating diarrhoea brought about by pathogenic  
CC bacteria and/or rotavirus. The carrier is a food composition selected  
CC from milk, yogurt, curd, cheese, fermented milks, milk based fermented  
CC products, ice-creams, fermented cereal based products, milk based  
CC powders, infant formula, pet food or a pharmaceutical composition  
CC selected from tablets, liquid bacterial suspensions, dried oral  
CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.  
CC (I) is useful in DNA arrays or chips to carry out analysis of the  
CC expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent  
CC Bifidobacterium related nucleotide sequences given in the sequence  
CC listing from the present invention but not mentioned further within the  
CC specification. N.B. The sequence data for this patent is not represented  
CC in the printed specification but is based on sequence information  
CC supplied by the European Patent Office  
XX  
SQ Sequence 322 AA;  
Query Match 6.8%; Score 113; DB 5; Length 322;  
Best Local Similarity 24.7%; Pred. No. 0.015;  
Matches 47; Conservative 29; Mismatches 74; Indels 40; Gaps 7;  
Qy 10 WKATPLVLTAGLRLPGEKAQKLLQKVEVFKASFFLVGDDCVSIMNGDEGVSAWIT 69  
Db 78 YEITQLLWATSALREAP--NGNKVIHKEE-----LIGQG-VTVLSGTDEA-----R 122  
Qy 70 INFITGSLKTPGGSSVG---MLDGGSGTQIAFLPRVEGTQLQASPPGYLTALRMFNRYK 126  
Db 123 LTFL--AARRWYGWDAGRLLVLDIGGSGLEVANGSDEEPTVALSAPA----- 167  
Qy 127 LYSYSYLGGLGMSARLAILGGVEGQPAKDGKELVSPCLSPSKGEWHAETVTVRVSGQKA 186  
Db 168 -----GAGRVTFEFLPSGNATPDELEDVKNVRKILEPMVKVFPQSKPHNHAVTGSKT 220  
Qy 187 AASLHELCAA 196  
Db 221 FRSLARLQGA 230  
RESULT 10  
AAM14306  
ID AAM14306 standard; protein; 45 AA.  
XX  
AC AAM14306;  
XX  
DT 12-OCT-2001 (first entry)  
XX  
DE Peptide #740 encoded by probe for measuring cervical gene expression.  
XX  
KW Probe; human; microarray; gene expression; cervical epithelial cell;  
KW cervical cancer.  
XX  
OS Homo sapiens.  
XX  
PN WO200157279-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000670.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488901/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human cervical epithelial cells.  
XX  
XX Claim 27; SEQ ID NO 19132; 487bp; English.  
XX  
XX The present invention relates to human single exon nucleic acid probes  
CC (SENSE: see AAI10068-AA128459). The present sequence is a peptide encoded  
CC by one such probe. The SENPs are derived from human Hela cells. The SENPs  
CC can be used to produce a single exon microarray, which can be used for  
CC measuring human gene expression in a sample derived from human cervical  
CC epithelial cells. By measuring gene expression, the probes are therefore  
CC useful in grading and/or staging of diseases of the cervix, notably  
CC cervical cancer. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 46 AA;  
SQ  
Query Match 6.3%; Score 104; DB 4; Length 46;  
Best Local Similarity 47.6%; Pred. No. 0.0068;  
Matches 20; Conservative 7; Mismatches 15; Indels 0; Gaps 0;  
QY 192 ELCAARYSEVLQNRHRTSEVKHVDYAFSYDYDLAAGVLI 233  
DB 5 EPCYAEVLRVVRKGLKHQPEVQSGSFYAFSYDRAVDITMI 46  
RESULT 11  
ID ABB33253 standard; peptide; 46 AA.  
XX  
XX ABB33253;  
XX  
XX 04-FEB-2002 (first entry)  
XX  
XX Peptide #759 encoded by human foetal liver single exon probe.  
XX  
XX Human; foetal liver; gene expression; single exon nucleic acid probe.  
XX  
XX Homo sapiens.  
XX  
XX WO200157277-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000669.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
XX  
XX 26-MAY-2000; 2000US-0207456P.  
XX  
XX 30-JUN-2000; 2000US-00608408.  
XX  
XX 03-AUG-2000; 2000US-00632366.  
XX  
XX 21-SEP-2000; 2000US-0234687P.  
XX  
XX 27-SEP-2000; 2000US-0236359P.  
XX  
XX 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-483447/52.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human fetal liver.  
XX

PS Claim 27; SEQ ID NO 25888; 639pp + Sequence listing; English.  
XX  
XX The invention relates to a single exon nucleic acid probe for measuring  
CC human gene expression in a sample derived from human foetal liver. The  
CC single exon nucleic acid probes may be used for predicting, measuring and  
CC displaying gene expression in samples derived from human fetal liver. The  
CC present sequence is a peptide encoded by a single exon nucleic acid probe  
CC of the invention. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 46 AA;  
SQ  
Query Match 6.3%; Score 104; DB 4; Length 46;  
Best Local Similarity 47.6%; Pred. No. 0.0068;  
Matches 20; Conservative 7; Mismatches 15; Indels 0; Gaps 0;  
QY 192 ELCAARYSEVLQNRHRTSEVKHVDYAFSYDYDLAAGVLI 233  
DB 5 EPCYAEVLRVVRKGLKHQPEVQSGSFYAFSYDRAVDITMI 46  
RESULT 12  
ID AAM26716 standard; protein; 46 AA.  
XX  
XX AAM26716;  
XX  
XX 17-OCT-2001 (first entry)  
XX  
XX Peptide #753 encoded by probe for measuring placental gene expression.  
XX  
XX Probe; microarray; human; placenta; antenatal diagnosis;  
XX genetic disorder.  
XX  
XX Homo sapiens.  
XX  
XX WO200157272-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000663.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
XX  
XX 26-MAY-2000; 2000US-0207456P.  
XX  
XX 30-JUN-2000; 2000US-00608408.  
XX  
XX 03-AUG-2000; 2000US-00632366.  
XX  
XX 21-SEP-2000; 2000US-0234687P.  
XX  
XX 27-SEP-2000; 2000US-0236359P.  
XX  
XX 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488897/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human placenta.  
XX  
XX Claim 27; SEQ ID NO 26985; 654pp; English.  
XX  
XX The present invention relates to single exon nucleic acid probes (SENPs:  
CC see AAI3315-AA157546). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for producing a microarray for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from human placenta. The probes are useful for antenatal diagnosis of  
CC human genetic disorders  
XX  
XX Sequence 46 AA;  
SQ  
Query Match 6.3%; Score 104; DB 4; Length 46;  
Best Local Similarity 47.6%; Pred. No. 0.0068;



Matches 20; Conservative 7; Mismatches 15; Indels 0; Gaps 0;

Qy 192 ELCAARVSEVLQNRVHRTTEVKHVDYAFSYDYDLAAGVGLI 233  
Db 5 EPCVAEVLVRVGRGLHQPEEVQSGSFYAFSYDYDLAAGVGLI 46

RESULT 13  
ABB28079  
ID ABB28079 standard; peptide; 46 AA.  
XX  
AC ABB28079;  
XX  
DT 01-FEB-2002 (first entry)  
XX  
DE Human peptide #730 encoded by breast cell single exon nucleic acid probe.  
XX  
KW Human; microarray; single exon probe; gene expression; breast; disease;  
KW cancer.  
XX  
OS Homo sapiens.  
XX  
FN WO200157271-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000662.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-496933/54.  
XX  
PT New spatially-addressable set of single exon nucleic acid probes, useful  
PT for measuring gene expression in sample derived from human breast,  
PT comprises number of single exon nucleic acid probes.  
XX  
PS Claim 27; SEQ ID NO 11047; 327pp + Sequence Listing; English.  
XX  
CC The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human breast and Br 474 cells. The method involves contacting the  
CC probes with a collection of detectably labelled nucleic acids derived  
CC from mRNA of human breast, and then measuring the label bound to each  
CC probe of the microarray. The probes are useful for verifying the  
CC expression of regions of genomic DNA predicted to encode proteins. They  
CC are useful for gene discovery, and for determining predisposition and/or  
CC prognosing breast disease. Gene expression analysis is useful for  
CC assessing the toxicity of chemical agents on cells. The microarray of  
CC this invention presents a far greater diversity of probes for measuring  
CC gene expression, with far less bias than expressed sequence tag  
CC microarrays. The method is suitable for rapid production of functional  
CC information from genomic sequence. The present sequence is a peptide  
CC encoded by a single exon nucleic acid probe of the invention. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 46 AA;

Query Match 6.3%; Score 104; DB 4; Length 46;  
Best Local Similarity 47.6%; Pred. No. 0.0068;  
Matches 20; Conservative 7; Mismatches 15; Indels 0; Gaps 0;

Qy 192 ELCAARVSEVLQNRVHRTTEVKHVDYAFSYDYDLAAGVGLI 233  
Db 5 EPCVAEVLVRVGRGLHQPEEVQSGSFYAFSYDYDLAAGVGLI 46

RESULT 14  
ABB18715  
ID ABB18715 standard; protein; 46 AA.  
XX  
AC ABB18715;  
XX  
DT 23-JAN-2002 (first entry)  
XX  
DE Protein #714 encoded by probe for measuring heart cell gene expression.  
XX  
KW Human; gene expression; heart; microarray; vascular system;  
KW cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease.  
XX  
OS Homo sapiens.  
XX  
FN WO200157274-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000666.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-488899/53.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
PT hearts.  
XX  
PS Claim 15; SEQ ID NO 20485; 530pp; English.  
XX  
CC The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart (see  
CC AB21535-ABA41305). The present sequence is a protein encoded by one such  
CC probe. The probes may be used for predicting, measuring and displaying  
CC gene expression in samples derived from the human heart via microarrays.  
CC By measuring gene expression, the probes are useful for predicting,  
CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
CC human heart and vascular system e.g. cardiovascular disease,  
CC hypertension, cardiac arrhythmias and congenital heart disease. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 46 AA;

Query Match 6.3%; Score 104; DB 4; Length 46;  
Best Local Similarity 47.6%; Pred. No. 0.0068;  
Matches 20; Conservative 7; Mismatches 15; Indels 0; Gaps 0;

Qy 192 ELCAARVSEVLQNRVHRTTEVKHVDYAFSYDYDLAAGVGLI 233  
Db 5 EPCVAEVLVRVGRGLHQPEEVQSGSFYAFSYDYDLAAGVGLI 46

RESULT 15  
AAM66436  
ID AAM66436 standard; protein; 46 AA.



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 10:10:12 ; Search time 48 Seconds  
(without alignments)  
1887.314 Million cell updates/sec

Title: US-09-905-589a-2\_COPY\_135\_456  
Perfect score: 1659  
Sequence: 1 AKQIPDFMKATPLVLRAT.....ALGAIPIHYIDSLNRKSPAS 322

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 830003

Minimum DB seq length: 0  
Maximum DB seq length: 322

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTIS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	237	14.3	278	US-10-425-114-71791	Sequence 71791, A
2	184.5	11.1	233	US-10-425-114-39418	Sequence 39418, A
3	175.5	10.6	199	US-10-424-599-219735	Sequence 219735, A
4	168.5	10.2	106	US-10-424-599-209550	Sequence 209550, A
5	157.5	9.5	288	US-10-424-599-262339	Sequence 262339, A
6	151.5	9.1	206	US-09-833-245-870	Sequence 870, App
7	144.5	8.7	219	US-10-424-599-266621	Sequence 266621, A
8	126.5	7.6	119	US-10-424-599-160616	Sequence 160616, A
9	124.5	7.5	112	US-10-425-114-69638	Sequence 49638, A
10	120.5	7.3	135	US-10-425-114-52324	Sequence 52324, A
11	113	6.8	186	US-10-424-599-195675	Sequence 195675, A
12	110.5	6.7	141	US-10-425-114-43552	Sequence 43552, A
13	110.5	6.7	155	US-10-424-599-150130	Sequence 150130, A
14	109.5	6.6	220	US-10-002-631C-276	Sequence 276, App
15	104	6.3	46	US-09-864-761-34013	Sequence 34013, A

15	95.5	5.8	54	12	US-10-424-599-260917	Sequence 260917, A
16	93	5.6	157	12	US-10-424-599-143499	Sequence 143499, A
17	86	5.2	280	12	US-10-156-761-9488	Sequence 9488, App
18	85.5	5.2	236	12	US-10-424-599-227975	Sequence 227975, A
19	85.5	5.1	293	15	US-10-369-493-2970	Sequence 2970, App
20	84	5.0	224	14	US-10-156-761-12887	Sequence 12887, A
21	83.5	5.0	167	12	US-10-424-599-281704	Sequence 281704, A
22	83	5.0	290	16	US-10-287-226-280	Sequence 280, App
23	83	5.0	192	12	US-10-363-616-469	Sequence 469, App
24	82.5	5.0	198	15	US-10-264-049-2947	Sequence 2947, App
25	82.5	5.0	228	12	US-10-425-114-39004	Sequence 39004, A
26	82.5	4.9	258	15	US-10-374-780A-1409	Sequence 1409, App
27	82	4.9	274	14	US-10-403-105-8	Sequence 8, App11
28	81.5	4.9	299	15	US-10-104-004-2792	Sequence 158177, App
29	81.5	4.9	150	12	US-10-424-599-158177	Sequence 158177, App
30	81	4.9	280	14	US-10-156-761-9491	Sequence 9491, App
31	80	4.8	274	14	US-10-336-324-5	Sequence 5, App11
32	79.5	4.8	301	12	US-10-425-114-36727	Sequence 36727, A
33	77	4.6	209	15	US-10-369-493-14007	Sequence 14007, A
34	76.5	4.6	252	10	US-09-984-130-146	Sequence 146, App
35	76.5	4.6	252	10	US-09-836-353A-146	Sequence 146, App
36	76.5	4.6	287	12	US-10-282-122A-61268	Sequence 61268, App
37	76.5	4.6	260	14	US-10-288-996-3	Sequence 3, App11
38	76	4.6	210	14	US-10-424-599-211969	Sequence 211969, App
39	76	4.6	273	16	US-10-389-566-477	Sequence 477, App
40	76	4.6	215	15	US-10-369-493-4676	Sequence 4676, App
41	75.5	4.6	215	15	US-10-369-493-7435	Sequence 7435, App
42	75.5	4.6	286	9	US-09-731-872-287	Sequence 287, App
43	75.5	4.6	286	9	US-09-876-997-287	Sequence 287, App
44	75.5	4.6	286	9	US-10-425-114-65997	Sequence 65997, A
45	75	4.5	244	12	US-10-425-114-65997	Sequence 65997, A

## ALIGNMENTS

RESULT 1  
US-10-425-114-71791  
; Sequence 71791, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jindong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(5313) B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 71791  
; LENGTH: 278  
; TYPE: PRT  
; ORGANISM: Zea mays subsp. mexicana  
; FEATURE:  
; OTHER INFORMATION: Clone ID: UC-ZMROTEOSINTE009H08\_FLI.pep  
US-10-425-114-71791  
Query Match 14.3%; Score 237; DB 12; Length 278;  
Best Local Similarity 25.8%; Pred. No. 2.5e-16;  
Matches 73; Conservative 51; Mismatches 109; Indels 50; Gaps 11;  
QY 68 ITTNFTGSLKTPGGSSVGMKLDIGGSGTQIAFPVREGTLQASPPG-----YLTALAM 120  
DB 1 VALNYLLDKLGGDYSQTVAVLDMGGSSVGMAY--ANSAVAASAPAPPHGEDPYVTRYL 58  
QY 121 FNTYLYGYSYSYGLGIMAPRLAIIIGVGAGPAKDKELVSPCLSPFGMEHAEVTR 180  
DB 59 KGDVNIYTHSYLHYAPASRAEIL-----KAKGDP--FSSCMKRGFGQTYNEQVD 110  
QY 181 VSGQKAASLHEICAAVSEVLQNRVHRFEVYKHDF-----YAFSYYPD 225

Db 111 ATAPEGA-VYKCRBEIGAL--NLNAPQCMNCTFNWGGGAGGQSIYVASSFYF 167  
Qy 226 LAAGVGLIDAE-KGGSLVGDFFIAAKYVCRTELOPO-----SSPFCMDLTY-V 274  
Db 168 VASEVGIVDGNAPSGNTPGAFGSAEAKACQMSVEEAKIEYPNVNDVDFYLCMDLAYQY 227  
Qy 275 SLLQBEFGPRSKVLKTRKIDN---VETSWALGAIFHYIDS 313  
Db 228 TLLVDGEGVETKEITVVDKVKGEYVEAANPLGSAIEAVSS 270  
RESULT 2  
US-10-425-114-39418  
; Sequence 39418, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 39418  
; LENGTH: 233  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700555327\_FLI.pep  
US-10-425-114-39418  
Query Match 11.1%; Score 184.5; DB 12; Length 233;  
Best Local Similarity 27.5%; Pred. No. 7.3e-11;  
Matches 64; Conservative 33; Mismatches 91; Indels 45; Gaps 8;  
Qy 114 YLTALRMFNRTYKLYSYLGLGLMSARLAILGGVEGQPAKDGKELVSPCLSPFKGEWE 173  
Db 4 YIKTLVNGKTYDIYVSHLHFHFGKEASRAEML-KVTGDSA-----NPCILAGYNGTYT 55  
Qy 174 HAEVTVRSQGAASALHELCAARVSEVLQNRVHRTVEEVGVD----- 216  
Db 56 YSGVKY-----KALASTSGSNFCKREVALKALKVNEPCPHQNTFGIWNNGGGSGQKV 110  
Qy 217 FYAPSVYVYDLAAGVGLIDAEKGGSLV-GDFEIAAKYVCRTELOPOSS-----PF 266  
Db 111 LYVTSFYLVQIGIADAKTSKYVPAEFKAARACQVKEFDAQSTYPLMMDALPY 170  
Qy 267 SCMDLTY-VSLLQBEFGPRSKVLKTRKIDN---VETSWALGAIFHYIDS 314  
Db 171 ICMDDITYQLTLVDGFLDEWKEIIVANEIEYQCALVEGAWPLGSAIEAIISS 223  
RESULT 3  
US-10-424-599-219735  
; Sequence 219735, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 219735  
; LENGTH: 199  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_40449C.1.pep  
US-10-424-599-219735  
Query Match 10.6%; Score 175.5; DB 12; Length 199;  
Best Local Similarity 34.9%; Pred. No. 5.2e-10;  
Matches 53; Conservative 21; Mismatches 59; Indels 19; Gaps 5;  
Qy 1 AKQDIPDFWKATPLVLKATAGLRLPGEKAQKLLQKVEFK-ASPFLVGDDCVSIMNG 59  
Db 53 AQKQIPKTSKHTTSLFLYATAGVRRRLPDDDSKMLLDNAWSFLKSSSPFVCKRDVWKIISG 112  
Qy 60 TDEGVSAWITINFLTGSILKT-PGGSSVGMGLDLGGSGTQIAFLPRVEGT-----LQ 108  
Db 113 TEAYLGMIALNYDSGILGVKPKETTYGALDLGGSSLOVTF-----EGNRNKEQQNQQLLN 168  
Qy 109 ASPFGYLTALRMFNRTYKLYSYLGLGLMSA 140  
Db 169 SETRLY---VRIGSMNHHLTGYSLAGYGLNEA 197  
RESULT 4  
US-10-424-599-209550  
; Sequence 209550, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 209550  
; LENGTH: 106  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_31250C.1.pep  
US-10-424-599-209550  
Query Match 10.2%; Score 168.5; DB 12; Length 106;  
Best Local Similarity 38.5%; Pred. No. 1.1e-09;  
Matches 35; Conservative 22; Mismatches 33; Indels 1; Gaps 1;  
Qy 1 AKQDIPDFWKATPLVLKATAGLRLPGEKAQKLLQKVEFKASPFILVG-DGVSIMNG 59  
Db 16 AESVVPRFRPKTPVRVGATAGLRALEGDASGRILQAVRDLLKORSTLKSEPDVAVLDG 75  
Qy 60 TDEGVSAWITINFLTGSILKTPEGSSVGMGLDL 90  
Db 76 TQEGAFQWVTNYLLGLKGLKDFSETVGVDL 106  
RESULT 5  
US-10-424-599-262339  
; Sequence 262339, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B

CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 262339  
LENGTH: 206  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(206)  
OTHER INFORMATION: unsure at all Xaa locations  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_78914C.1.pep  
US-10-424-599-262339

Query Match 9.5%; Score 157.5; DB 12; Length 206;  
Best Local Similarity 42.3%; Pred. No. 4,5e-08;  
Matches 33; Conservative 17; Mismatches 27; Indels 1; Gaps 1;

QY 1 AKODIPDFWKATPVLKATAGLRLPGEKAQKLLQKVEVK-ASFLVGDGVSIANG 59  
120 AASVPEEDYPTTPVGLGATAGRLLEGDSERILQAVADLKNSTNTQPDVAIIDD 179  
DB 60 TDEGVSAMITINFL 77  
180 TQEGSYLWVITNYLKL 197

RESULT 6  
US-09-833-245-870  
Sequence 870, Application US/09833245  
Publication No. US20040010134A1  
GENERAL INFORMATION:  
APPLICANT: Human Genome Sciences, Inc.  
TITLE OF INVENTION: Albumin Fusion Proteins  
FILE REFERENCE: PFS46PCT  
CURRENT APPLICATION NUMBER: US/09/833,245  
CURRENT FILING DATE: 2001-04-12  
PRIOR APPLICATION NUMBER: 60/229, 358  
PRIOR FILING DATE: 2000-04-12  
PRIOR APPLICATION NUMBER: 60/256, 931  
PRIOR FILING DATE: 2000-12-21  
PRIOR APPLICATION NUMBER: 60/199, 384  
PRIOR FILING DATE: 2000-04-25  
NUMBER OF SEQ ID NOS: 2267  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 870  
LENGTH: 288  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (67)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (99)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (230)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (263)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (264)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (270)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (275)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-833-245-870

Query Match 9.1%; Score 151.5; DB 11; Length 288;  
Best Local Similarity 35.1%; Pred. No. 3.3e-07;  
Matches 40; Conservative 14; Mismatches 45; Indels 15; Gaps 2;

QY 1 AKODIPDFWKATPVLKATAGLRLPGEKAQKLLQKVEVKASFLVGDGVSIANG 59  
DB 155 AAHVVPKKEKEMETPLICTAGRLPERKQLALADLVLDLPEDFLFSQSAEYISG 214  
QY 60 TDEGVSAMITINFL 77  
DB 215 KQESVAMIGINFLVKRFDHEDSDAEATQELAAARRRIVGILMDGAXXQILAY 268

RESULT 7  
US-10-424-599-266621  
Sequence 266621, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei

TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 266621  
LENGTH: 219  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (219)  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_82780C.1.pep  
US-10-424-599-266621

Query Match 8.7%; Score 144.5; DB 12; Length 219;  
Best Local Similarity 25.9%; Pred. No. 1.2e-06;  
Matches 57; Conservative 40; Mismatches 80; Indels 43; Gaps 10;

QY 135 LGLMSARLALIL-----GVEGQPA---KQKEVSPCLSPSFKGEMHAEVTVYVSG 183  
DB 1 MGLLRGMSVLVGHGVHGLAARVPLSSVALSPVGLPGAGSYVYGKSPFASS 60  
QY 184 QKAASLHE-----LCARVSEVLQNRVHRT-----EYKAVDFYASYYDLAA 228  
DB 61 GSSGASLNECKSVLRALRKVNESTCTHMKCTPGGIWNGGGGQKXLL--FVASFFDRAA 118  
QY 229 GVGILDEKGSILV-VQDFEIAKVCRTLETQPOSS-----PSCMDLTLY-VSL 277  
DB 119 EAGPADNLVALVPRADFDAAKQACQTLLENAKSTFPVVDGNNLPYLCMDLIQYITLL 178  
QY 278 LQEFGEPRSKVLTXTKI--DN--VETSWALGAIFHYIDS 313  
DB 179 VDFGIYPMQBITLVKKYKDDALVEAAWPLGSAIAVSS 218

RESULT 8  
US-10-424-599-160816  
Sequence 160816, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684

```
; SEQ ID NO 160816
; LENGTH: 119
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_116234C.1.pep
US-10-424-599-160816

Query Match
Best Local Similarity 7.6%; Score 126.5; DB 12; Length 119;
Matches 27; Conservative 22; Mismatches 43; Indels 5; Gaps 1;

QY 40 EVKASPLVGDCCVSWTGTGSGVATITNLTGSLKTPGSSVGMGLDGGSGTQIAF 99
Db 1 QAKVSGFLFMDWARVISGEQGISWAVATNLTGSLKTPGSSVGMGLDGGSGTQIAF 99
QY 100 LPRVEGTQASPPGYLTALRMFNRTYKLYSYSLGLG 136
Db 59 ---TSAKLNADIAQSLHTIRLSGVMYNLYTRSLPQLG 92

RESULT 9
US-10-425-114-49638
; Sequence 49638, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 49638
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700751504_FLI.pep
US-10-425-114-49638

Query Match
Best Local Similarity 7.5%; Score 124.5; DB 12; Length 171;
Matches 35; Conservative 21; Mismatches 42; Indels 17; Gaps 5;

QY 216 DFYAFSYYYDLAAGVLIDAEKGGSLV-VGDFFETAAKYVCTLTETOPQSS----- 264
Db 48 NFFIASPFVEADGAFVDPNAPNAKVRPVDFFENAAKVACT-ELKDKLSIFPRVKGDV 106
QY 265 PFCMDLTY-VSLLLQERGFPRSKVLKTRKIDN---VETSWALGAIFHYDLS 314
Db 107 PYICLDLVYETLLVDGFDIDPQOEITLVKQVEYQDSLVAEAAWPLGSAIEAISL 161

RESULT 10
US-10-425-114-52324
; Sequence 52324, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
```

```
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 52324
; LENGTH: 135
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700788139_FLI.pep
US-10-425-114-52324

Query Match
Best Local Similarity 7.3%; Score 120.5; DB 12; Length 135;
Matches 35; Conservative 21; Mismatches 42; Indels 15; Gaps 5;

QY 216 DFYAFSYYYDLAAGVLIDAEKGGSLV-VGDFFETAAKYVCTLTETOPQSS-----P 265
Db 22 NLFVASPFEDRAAGFADPNLPAIVRPADFDAAKQACQTKLENAKSTFFRVDEGNLP 81
QY 266 FSCMDLTY-VSLLLQERGFPRSKVLKTRKIDN---VETSWALGAIFHYDLS 313
Db 82 YLCMDLIYQYTLVDGFGIYFWQETILVKKYDDALVEAAWPLGSAIEAVSS 134

RESULT 11
US-10-424-599-195675
; Sequence 195675, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 195675
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(186)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_18720C.1.pep
US-10-424-599-195675

Query Match
Best Local Similarity 6.8%; Score 113; DB 12; Length 186;
Matches 49; Conservative 23; Mismatches 61; Indels 58; Gaps 10;

QY 148 VEGQPAKDGKELV-----SPCLSPSPKGEWEAEVTVYRVSGQKAAASLHELCAARV 198
Db 20 IGGSKXDECRVVLQALKLNESCCHONCTFGINDGG----RGSGQK----- 62
QY 199 SEVLQNRVHRTVEEVKHVDYFAFSYYIDLAAGVLIDAEKGGSLV-VGDFFETAAKYVCT- 256
Db 63 -----ILFGTSSFFY-LPTBIGIIDLNKPNKIHVPDLIEAKGACETK 105
QY 257 LE-----TOPQSS-----PFCMDLTY-VSLLLQERGFPRSKVLKTRKIDN---VETSWA 303
Db 106 LEDAKSTYPNTLAEDRLFPVCLDIAYQYALYTDGFDLPWQBITVANELEYQDALVEAAW 165
QY 304 LGAIFHYDLS 314
Db 166 LGTAIEAISL 176

RESULT 12
```



```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34013
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005480.3
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.8
; OTHER INFORMATION: EST HUMAN HIT: BE780916.1, EVALUE 4.00e-20
; OTHER INFORMATION: EST_HUMAN HIT: BE407814.1, EVALUE 2.00e-04
US-09-864-761-34013
```

```
Query Match          6.3%  Score 104; DB 9; Length 46;
Best Local Similarity 47.6%  Pred. No. 0.0023;
Matches 20; Conservative 7; Mismatches 15; Indels 0; Gaps 0;

Qy 192 ELCAARVSEVLQNRVHRTTEVKHVDYAFSYYYDLAAGVGLI 233
   |||||:::|||||
Db 5 EPCVAEVLVRGKLGHPVEVQSGFYAFSYYYDRAVDTDMI 46
```

Search completed: June 8, 2004, 10:16:01  
Job time : 49 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: June 8, 2004, 10:07:32 ; Search time 23 Seconds  
(without alignments)  
722.764 Million cell updates/sec

Title: US-09-905-589A-2\_COPY\_135\_456  
Perfect score: 1659  
Sequence: 1 AKGDIPDFMKATPLVLKAT.....ALGAFHYIDSLNRQSPAS 322

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 335687

Minimum DB seq length: 0  
Maximum DB seq length: 322

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCUTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	121.5	7.3	148	4	US-09-240-639-17
2	121	7.3	150	4	US-09-240-639-16
3	111.5	6.7	153	4	US-09-240-639-13
4	110.5	6.7	154	4	US-09-240-639-14
5	109.5	6.6	153	4	US-09-240-639-15
6	81.5	4.9	274	4	US-09-196-281-8
7	81	4.9	199	1	US-08-264-534-4
8	81	4.9	199	1	US-08-465-500-4
9	81	4.9	199	2	US-08-346-126-4
10	81	4.9	199	2	US-08-346-126-4
11	81	4.9	199	3	US-08-893-828-4
12	81	4.9	293	1	US-08-264-534-9
13	81	4.9	293	1	US-08-083-590A-4
14	81	4.9	293	1	US-08-465-500-9
15	81	4.9	293	2	US-08-346-126-9
16	81	4.9	293	2	US-08-346-126-9
17	81	4.9	293	3	US-08-533-384-4
18	81	4.9	293	3	US-08-893-828-9
19	79.5	4.8	274	4	US-08-401-575A-1
20	79.5	4.8	274	4	US-09-512-251A-5
21	79.5	4.8	274	4	US-09-515-150A-5
22	79	4.8	224	4	US-09-489-039A-9852
23	76	4.6	210	4	US-09-439-226-3
24	76	4.6	285	4	US-09-328-352-5143
25	74.5	4.5	274	4	US-09-104-623A-3
26	74.5	4.5	274	4	US-09-019-532-3
27	74.5	4.5	274	4	US-09-338-746-3

28	74	4.5	291	4	US-09-489-039A-B172	Sequence 8172, Ap
29	73.5	4.4	270	2	US-08-652-507-2	Sequence 2, Appli
30	73.5	4.4	283	3	US-09-036-987A-13	Sequence 13, Appli
31	73.5	4.4	283	3	US-09-370-700-13	Sequence 13, Appli
32	73.5	4.4	283	4	US-09-603-207-13	Sequence 13, Appli
33	73	4.4	305	4	US-09-252-991A-19225	Sequence 19225, A
34	72	4.3	161	4	US-09-328-352-5028	Sequence 5028, Ap
35	72	4.3	313	4	US-09-148-545-236	Sequence 236, App
36	71.5	4.3	222	4	US-09-439-226-4	Sequence 4, Appli
37	71.5	4.3	278	3	US-08-491-954-2	Sequence 2, Appli
38	71.5	4.3	318	4	US-09-252-991A-19707	Sequence 19707, A
39	71	4.3	306	4	US-09-252-991A-29549	Sequence 29549, A
40	70.5	4.2	125	4	US-09-621-976-6812	Sequence 6812, Ap
41	70.5	4.2	216	4	US-09-489-039A-10545	Sequence 10545, A
42	70.5	4.2	284	4	US-09-914-259-52	Sequence 52, Appli
43	70.5	4.2	287	4	US-09-543-681A-5639	Sequence 5639, Ap
44	70.5	4.2	307	4	US-09-489-039A-8048	Sequence 8048, Ap
45	70	4.2	147	3	US-08-751-359-18	Sequence 18, Appli

## ALIGNMENTS

```
RESULT 1
US-09-240-639-17
; Sequence 17, Application US/09240639
; Patent No. 6350447
; GENERAL INFORMATION:
; APPLICANT: Chadwick, Brian Paul
; APPLICANT: Fritschaut, Anna-Maria
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO CD39-LIKE
; FILE REFERENCE: 9598-066
; CURRENT APPLICATION NUMBER: US/09/240,639
; CURRENT FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-240-639-17

Query Match      7.3%, Score 121.5, DB 4, Length 148;
Best Local Similarity 32.9%, Pred. No. 4.1e-05;
Matches 25; Conservative 22; Mismatches 28; Indels 1; Gaps 1;

QY      1 AKGDIPDFMKATPLVLKATGRLPGKAKLQKYEVEFKASPL-VGDDCVSINMG 59
Db      70 AERHPIYEQGTDLIFRTAGKRLLPENAKAIITKNLQNGKSYTALRVSDSNIRIIDG 129
QY      60 TDEGVSAWITNPLTG 75
Db      130 AWEGIYSWIAVNYILG 145

RESULT 2
US-09-240-639-16
; Sequence 16, Application US/09240639
; Patent No. 6350447
; GENERAL INFORMATION:
; APPLICANT: Chadwick, Brian Paul
; APPLICANT: Fritschaut, Anna-Maria
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO CD39-LIKE
; FILE REFERENCE: 9598-066
; CURRENT APPLICATION NUMBER: US/09/240,639
; CURRENT FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 150
; TYPE: PRT
```

```
; ORGANISM: Gallus gallus
US-09-240-639-16

Query Match      7.3%; Score 121; DB 4; Length 150;
Best Local Similarity 41.3%; Pred. No. 4.8e-05;
Matches 31; Conservative 10; Mismatches 30; Indels 4; Gaps 2;

QY 1 AKQDIPDFWKATPLVLKATAGLRLPCE---PGEKAQKLQKVEVFKASPFVLVGGDDCVSINMGTD 58
Db 66 AMEDVPKEKHADTPVYLKATAGMELLTIADPPSQTCLSAVMATLKSYPDFGG--AKILS 123

QY 59 GTDEGVSANITINFL 73
Db 124 GEEGVFGWITANYL 138

RESULT 3
US-09-240-639-13
; Sequence 13, Application US/09240639
; Patent No. 6350447
; GENERAL INFORMATION:
; APPLICANT: Chadwick, Brian Paul
; APPLICANT: Frischauf, Anna-Maria
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO CD39-LIKE
; TITLE OF INVENTION: POLYPEPTIDES AND NUCLEIC ACIDS
; FILE REFERENCE: 9598-066
; CURRENT APPLICATION NUMBER: US/09/240,639
; CURRENT FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-240-639-13

Query Match      6.7%; Score 111.5; DB 4; Length 153;
Best Local Similarity 40.0%; Pred. No. 0.00052;
Matches 34; Conservative 14; Mismatches 28; Indels 9; Gaps 4;

QY 1 AKQDIPDFWKATPLVLKATAGLRLPGEKAQKLQKVEVFKAS-----PFLVGGDDCVSI 56
Db 65 AREVIPSQHQETPPVYLKATAGMRLLRME--SESLADRVLDVWERSLSNYFF--DFQGARI 121

QY 57 MNGTDEGVSANITINFLTG--SLKT 79
Db 122 ITGQEGAYGWIINYLKGFESQKT 146

RESULT 4
US-09-240-639-14
; Sequence 14, Application US/09240639
; Patent No. 6350447
; GENERAL INFORMATION:
; APPLICANT: Chadwick, Brian Paul
; APPLICANT: Frischauf, Anna-Maria
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO CD39-LIKE
; TITLE OF INVENTION: POLYPEPTIDES AND NUCLEIC ACIDS
; FILE REFERENCE: 9598-066
; CURRENT APPLICATION NUMBER: US/09/240,639
; CURRENT FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-240-639-14

Query Match      6.7%; Score 110.5; DB 4; Length 154;
Best Local Similarity 40.5%; Pred. No. 0.00068;
Matches 30; Conservative 8; Mismatches 31; Indels 5; Gaps 2;
```

```
QY 5 IPDFDFWKATPLVLKATAGLRLPCE---KAQKLQKVEVFKASPFVLVGGDDCVSINMGTD 61
Db 70 IPASKQHQTTPVYLKATAGMELLRMESQSADEVLAASVRSLSKSYFF--DFQGAKITGOE 127

QY 62 EGVSANITINFLTG 75
Db 128 EGAYGWITINYLIG 141

RESULT 5
US-09-240-639-15
; Sequence 15, Application US/09240639
; Patent No. 6350447
; GENERAL INFORMATION:
; APPLICANT: Chadwick, Brian Paul
; APPLICANT: Frischauf, Anna-Maria
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO CD39-LIKE
; TITLE OF INVENTION: POLYPEPTIDES AND NUCLEIC ACIDS
; FILE REFERENCE: 9598-066
; CURRENT APPLICATION NUMBER: US/09/240,639
; CURRENT FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-240-639-15

Query Match      6.6%; Score 109.5; DB 4; Length 153;
Best Local Similarity 38.5%; Pred. No. 0.00086;
Matches 30; Conservative 10; Mismatches 29; Indels 9; Gaps 3;

QY 1 AKQDIPDFWKATPLVLKATAGLRLP---PGEKAQKLQKVEVFKASPFVLVGGDC--VS 55
Db 65 ALQDVPKEXRHAGTPLYLGATAGMELLNLTNPETASTSVIMAVTHLTQYPP----DFRGAR 120

QY 56 IMNGTDEGVSANITINFL 73
Db 121 ILSQEGVFGWVTANYL 138

RESULT 6
US-09-196-281-8
; Sequence 8, Application US/09196281A
; Patent No. 6605458
; GENERAL INFORMATION:
; APPLICANT: Hansen, Peter K.
; APPLICANT: Bauditz, Peter
; APPLICANT: Mikkelsen, Frank
; TITLE OF INVENTION: Protease Variants And Compositions
; FILE REFERENCE: 5435.200-US
; CURRENT APPLICATION NUMBER: US/09/196,281A
; CURRENT FILING DATE: 1998-11-19
; EARLIER APPLICATION NUMBER: 1332/97
; EARLIER FILING DATE: 1997-11-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 274
; TYPE: PRT
; ORGANISM: Bacillus
US-09-196-281-8

Query Match      4.9%; Score 81.5; DB 4; Length 274;
Best Local Similarity 22.0%; Pred. No. 2.3;
Matches 69; Conservative 50; Mismatches 118; Indels 77; Gaps 17;

QY 22 GIRLLPGEKAQKLQKVEVFKASPFVLVGGDDCVSINMGTDGVSANITINFLTGLSKTPG 81
Db 7 GIPLIKADKVG-----AQGYKANVKVGI-----IDTGAASVEAAAQHTDLKRVG 52

QY 82 GSSVGM-----LDLGGGSGTQ-----IAFLPRVEGTQLQSPFGYLTALRMFNRTYKLSYSY 132
```

Db 53 GASFEVSGSYNTDNGHGHVAGVAAALDNTTGVGVAPNVLVAIKLNS-----104  
QY 133 LGLGMSARLAILLGVEGPAKDKELVSPCL-SPSPFGMEHAEVTVRVSQKAAALH 191  
Db 105 -GSGTYS---AIVSIE-WATONGLDVIMSLGGS-----GSTAKQAVDKAYASGI 152  
QY 192 ELCAA--RVSEVLQNRVHRTVEVHRVDYFAFSYYDLAAGVGLDAEKG--SLVVG 245  
Db 153 VVVAAGNSGSSGQNTI-----GYPAKYDSVIAGVADSNKKNASFSVGA 200  
QY 246 FEIAKYCTLEFPOSPSPSCMDLTV-----SLLQFGRSKVLTETKID 296  
Db 201 LEVAPGV-SVYSTP-SNTYTSLNGTSMASPHVAGAAALILSK--YPTLASQVRNLS 256  
QY 297 NVETSMALGALFHY 310  
Db 257 STATN-IGDSFY 268

RESULT 7  
US-08-264-534-4  
Sequence 4, Application US/08264534  
Patent No. 5648464  
GENERAL INFORMATION:  
APPLICANT: Artavanis-Tsakonas, Spyridon et al.  
TITLE OF INVENTION: Human No. 5648464ch And Delta, Binding Domains  
TITLE OF INVENTION: In Toporythmic Proteins, And Methods Based Thereon  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/264,534  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mirock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-004  
TELEPHONE: 212 790-9090  
TELEFAX: 212 869864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 199 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-264-534-4

Query Match 4.9%; Score 81; DB 1; Length 199;  
Best Local Similarity 21.6%; Pred. No. 1.6;  
Matches 42; Conservative 21; Mismatches 65; Indels 66; Gaps 6;

QY 14 PLYKATAGRLIPGEKAKLQKVEVKASPLVGVDDCVSINAGTDEGVSAWTTINFL 73  
Db 26 PAELRAKTKIGSPCTTAFRL-----CLKEYOTTEGASISTGCSFG 67

QY 74 TGS�KTPGSSVGMULDGGSGTQIAFLPVEGTLOASPPGYTALRMFRRTKLYSYSL 133  
Db 68 NATTKILGSSSFLSPGVGATVLPFTFRWTSFTL-----ILQALDMVNTSY-----115  
QY 134 GGLGMSARLAILGVGPAKDKELVSPCLSPFG-----EWB-----HAEVTVR 180  
Db 116 -----PDAERLIE---ETSISGVILPSPFKKTLDIRNRARITR 152  
QY 181 VSGQKAAAGLHEL 194  
Db 153 VRVQCAVTVYNTTC 166

RESULT 8  
US-08-465-500-4  
Sequence 4, Application US/08465500  
Patent No. 5789195  
GENERAL INFORMATION:  
APPLICANT: Muskavitch, Marc A.T.  
APPLICANT: Fenon, Richard G.  
APPLICANT: Reday, Iarla  
APPLICANT: Blaumuller, Cristine M.  
APPLICANT: Shepard, Scott B.  
TITLE OF INVENTION: HUMAN NOTCH AND DELTA, BINDING DOMAINS  
TITLE OF INVENTION: IN TOPORYTHMIC PROTEINS, AND METHODS BASED THEREON  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465,500  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mirock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-034  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8664/9741  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 199 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-465-500-4

Query Match 4.9%; Score 81; DB 1; Length 199;  
Best Local Similarity 21.6%; Pred. No. 1.6;  
Matches 42; Conservative 21; Mismatches 65; Indels 66; Gaps 6;

QY 14 PLYKATAGRLIPGEKAKLQKVEVKASPLVGVDDCVSINAGTDEGVSAWTTINFL 73  
Db 26 PAELRAKTKIGSPCTTAFRL-----CLKEYOTTEGASISTGCSFG 67  
QY 74 TGS�KTPGSSVGMULDGGSGTQIAFLPVEGTLOASPPGYTALRMFRRTKLYSYSL 133  
Db 68 NATTKILGSSSFLSPGVGATVLPFTFRWTSFTL-----ILQALDMVNTSY-----115  
QY 134 GGLGMSARLAILGVGPAKDKELVSPCLSPFG-----EWB-----HAEVTVR 180

Db 116 -----PDAERLIE---ETSYSVILPSPWKTLDHIGNARITYR 152

QY 181 VSGKAAASLHELIC 194

Db 153 VRVQCAVYYNTTC 166

RESULT 9

US-08-346-126-4

; Sequence 4, Application US/08346126

; Patent No. 5849869

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, Spyridon et al.

; TITLE OF INVENTION: Human No. 5849869ch And Delta, Binding Domains

; TITLE OF INVENTION: In Toporythmic Proteins, And Methods Based Thereon

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/346,126

; FILING DATE:

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/791,923

; FILING DATE: 14-NOV-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Mirock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 7326-007

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212 790-9090

; TELEFAX: 212 8698864/9741

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: unknown

; MOLECULE TYPE: peptide

US-08-346-126-4

Query Match 4.9%; Score 81; DB 2; Length 199;

Best Local Similarity 21.6%; Pred. No. 1.6;

Matches 42; Conservative 21; Mismatches 65; Indels 66; Gaps 6;

QY 14 PLVLKATAGLLPGEKAQKLLQKVEFKASPLVGGDDCVSIMNGTDEGVSAWITINFL 73

Db 26 PAELRAITKICSPCTTAFL-----CLKEYQTTEQGASISTGCSFG 67

QY 74 TGSLLKTPGGSVGMLDLGGGSGTQIAFLPRVEGTLOASPPGYLTALRMFNRTYKLYSYL 133

Db 68 NATTKILGSSFVLSDPGVGAIVLFFTRWTKSFTL-----ILQALDMYNTSY-----115

QY 134 GLGLMSARLAILGGVEGPAKDGKELVSPCLSPSKFG-----EWE-----HAEVYR 180

Db 116 -----PDAERLIE---ETSYSVILPSPWKTLDHIGNARITYR 152

QY 181 VSGKAAASLHELIC 194

Db 153 VRVQCAVYYNTTC 166

RESULT 11

US-08-893-828-4

; Sequence 4, Application US/08893828

; Patent No. 6030922

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, Spyridon

; APPLICANT: Muskavitch, Marc A.T.

RESULT 10

US-08-346-128-4

; Sequence 4, Application US/08346128

; Patent No. 5856441

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, Spyridon et al.

; TITLE OF INVENTION: Human No. 5856441ch And Delta, Binding Domains

; TITLE OF INVENTION: In Toporythmic Proteins, And Methods Based Thereon

; NUMBER OF SEQUENCES: 37

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/346,128

; FILING DATE:

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/879,038

; FILING DATE: 30-APR-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Mirock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 7326-009

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212 790-9090

; TELEFAX: 212 8698864/9741

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 199 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: unknown

; MOLECULE TYPE: peptide

US-08-346-128-4

Query Match 4.9%; Score 81; DB 2; Length 199;

Best Local Similarity 21.6%; Pred. No. 1.6;

Matches 42; Conservative 21; Mismatches 65; Indels 66; Gaps 6;

QY 14 PLVLKATAGLLPGEKAQKLLQKVEFKASPLVGGDDCVSIMNGTDEGVSAWITINFL 73

Db 26 PAELRAITKICSPCTTAFL-----CLKEYQTTEQGASISTGCSFG 67

QY 74 TGSLLKTPGGSVGMLDLGGGSGTQIAFLPRVEGTLOASPPGYLTALRMFNRTYKLYSYL 133

Db 68 NATTKILGSSFVLSDPGVGAIVLFFTRWTKSFTL-----ILQALDMYNTSY-----115

QY 134 GLGLMSARLAILGGVEGPAKDGKELVSPCLSPSKFG-----EWE-----HAEVYR 180

Db 116 -----PDAERLIE---ETSYSVILPSPWKTLDHIGNARITYR 152

QY 181 VSGKAAASLHELIC 194

Db 153 VRVQCAVYYNTTC 166

RESULT 11

US-08-893-828-4

; Sequence 4, Application US/08893828

; Patent No. 6030922

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, Spyridon

; APPLICANT: Muskavitch, Marc A.T.

APPLICANT: Rehon, Richard G.  
APPLICANT: Rebay, Ilaria  
APPLICANT: Blumel, Cristine M.  
APPLICANT: Shepard, Scott B.  
TITLE OF INVENTION: HUMAN NOTCH AND DELTA, BINDING DOMAINS  
TITLE OF INVENTION: IN TOPOGRAPHIC PROTEINS, AND METHODS BASED THEREON  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESS: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/893,828  
FILING DATE: 11-JUL-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mastrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-050  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 199 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-893-828-4

Query Match 4.9%; Score 81; DB 3; Length 199;  
Best Local Similarity 21.6%; Pred. No. 1.6;  
Matches 42; Conservative 21; Mismatches 65; Indels 66; Gaps 6;

QY 14 PLVLTATGRLPGRKQKLVKASPLVGVDDCVSINMGDEGVSAITINFL 73  
DB 26 PAELRATITIGSPCTTAFRL-----CLKEYQTTEGASISTGCSFG 67  
QY 74 TGSLLTPGSSSVGMDDLGGSTQIAFLPRVEGTLOASPPGYLTALRMENRTYKLYSYSL 133  
DB 68 NATTKILGSSSVLSPGVGAIVLPFTFRWTKSFTL-----ILQALDMVNTSY----- 115  
QY 134 GUGMSARLAILGVGQPAKDGEVSPCLSPFGK-----EWE-----HAEVYR 180  
DB 116 -----PDAERLIE-----ETSYSGVILPSPEWKTLDHGRNARITR 152  
QY 181 VSGQKAASLHSLC 194  
DB 153 VRVQCAVITYNTTC 166

RESULT 12  
US-08-264-534-9  
Sequence 9, Application US/08264534  
Patent No. 5648464  
GENERAL INFORMATION:  
APPLICANT: Atavanis-Tsakonas, Spyridon et al.  
TITLE OF INVENTION: Human No. 5648464ch And Delta, Binding Domains  
TITLE OF INVENTION: In Topography Proteins, And Methods Based Thereon  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSES: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York

STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/264,534  
FILING DATE: 03-MAY-1991  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/695,189  
FILING DATE: 03-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Mastrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090  
TELEFAX: 212 8698864/9741  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 293 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-264-534-9

Query Match 4.9%; Score 81; DB 1; Length 293;  
Best Local Similarity 21.6%; Pred. No. 2.9;  
Matches 42; Conservative 21; Mismatches 65; Indels 66; Gaps 6;

QY 14 PLVLTATGRLPGRKQKLVKASPLVGVDDCVSINMGDEGVSAITINFL 73  
DB 106 PAELRATITIGSPCTTAFRL-----CLKEYQTTEGASISTGCSFG 147  
QY 74 TGSLLTPGSSSVGMDDLGGSTQIAFLPRVEGTLOASPPGYLTALRMENRTYKLYSYSL 133  
DB 148 NATTKILGSSSVLSPGVGAIVLPFTFRWTKSFTL-----ILQALDMVNTSY----- 195  
QY 134 GUGMSARLAILGVGQPAKDGEVSPCLSPFGK-----EWE-----HAEVYR 180  
DB 196 -----PDAERLIE-----ETSYSGVILPSPEWKTLDHGRNARITR 232  
QY 181 VSGQKAASLHSLC 194  
DB 233 VRVQCAVITYNTTC 246

RESULT 13  
US-08-083-590A-4  
Sequence 4, Application US/08083590A  
Patent No. 5786158  
GENERAL INFORMATION:  
APPLICANT: Atavanis-Tsakonas, S. et al.  
TITLE OF INVENTION: Therapeutic And Diagnostic Methods  
TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And  
TITLE OF INVENTION: Nucleic Acids  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSES: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS

/ SOFTWARE: PatentIn Release #1.0, Version #1.25

/ CURRENT APPLICATION DATA:

/ APPLICATION NUMBER: US/08/083,590A

/ FILING DATE: 25-JUN-1993

/ CLASSIFICATION: 435

/ ATTORNEY/AGENT INFORMATION:

/ NAME: Mierock, S. Leslie

/ REGISTRATION NUMBER: 18,872

/ REFERENCE/DOCKET NUMBER: 7326-015

/ TELECOMMUNICATION INFORMATION:

/ TELEPHONE: 212 790-9090

/ TELEFAX: 212 8698864/9741

/ TELEX: 66141 PENNIE

/ INFORMATION FOR SEQ ID NO: 4:

/ SEQUENCE CHARACTERISTICS:

/ LENGTH: 293 amino acids

/ TYPE: amino acid

/ TOPOLOGY: unknown

/ MOLECULE TYPE: protein

/ US-08-083-590A-4

Query Match 4.9%; Score 81; DB 1; Length 293;

Best Local Similarity 21.6%; Pred. No. 2.9; Mismatches 65; Indels 66; Gaps 6;

Matches 42; Conservative 21; Mismatches 65; Indels 66; Gaps 6;

QY 14 PLVLKATAGRLRLPGKAKLQKVKFKASFLVGGDDCVSIMNGTDEGVSAMITINFL 73

Db 106 PAELRATKIGCSPCTTAFRL-----CLKEYQTTEQASISTGCSFG 147

QY 74 TGS�KTPGSSVGMGLDGGSGTOIAFLPRVEGTLOASPPGYLTALRMFNRTYKLYSYL 133

Db 148 NATTKILGGSSFVLSDPGVAIVLPTFTFRWTKSFTL-----ILOALDMNTSY----- 195

QY 134 GLGLMSARLAILGGVEGQPAKGKELVSPCLSPSPFKG-----EWE-----HAEVTVYR 180

Db 196 -----PDAERLIE-----ETSYSVILPSPENKTLDHIGNRARIYR 232

QY 181 VSCQKAAASLHELIC 194

Db 233 VRVQCAVYYNTTC 246

RESULT 14

US-08-465-500-9

; Sequence 9, Application US/08465500

; Patent No. 5789195

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, Spyridon

; APPLICANT: Muskavitch, Marc A.T.

; APPLICANT: Feron, Richard G.

; APPLICANT: Rebay, Ilaria

; APPLICANT: Blaumueller, Cristine M.

; APPLICANT: Shepard, Scott B.

; TITLE OF INVENTION: HUMAN NOTCH AND DELTA, BINDING DOMAINS

; TITLE OF INVENTION: IN TOPORYTHMIC PROTEINS, AND METHODS BASED THEREON

; NUMBER OF SEQUENCES: 34

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PENNIE &amp; EDMONDS

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: NY

; COUNTRY: USA

; ZIP: 10036-2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/465,500

; FILING DATE: 05-JUN-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

/ NAME: Mierock, S. Leslie

/ REGISTRATION NUMBER: 18,872

/ REFERENCE/DOCKET NUMBER: 7326-034

/ TELECOMMUNICATION INFORMATION:

/ TELEPHONE: (212) 790-9090

/ TELEFAX: (212) 869-8864/9741

/ INFORMATION FOR SEQ ID NO: 9:

/ SEQUENCE CHARACTERISTICS:

/ LENGTH: 293 amino acids

/ TYPE: amino acid

/ TOPOLOGY: unknown

/ MOLECULE TYPE: protein

/ US-08-465-500-9

Query Match 4.9%; Score 81; DB 1; Length 293;

Best Local Similarity 21.6%; Pred. No. 2.9;

Matches 42; Conservative 21; Mismatches 65; Indels 66; Gaps 6;

QY 14 PLVLKATAGRLRLPGKAKLQKVKFKASFLVGGDDCVSIMNGTDEGVSAMITINFL 73

Db 106 PAELRATKIGCSPCTTAFRL-----CLKEYQTTEQASISTGCSFG 147

QY 74 TGS�KTPGSSVGMGLDGGSGTOIAFLPRVEGTLOASPPGYLTALRMFNRTYKLYSYL 133

Db 148 NATTKILGGSSFVLSDPGVAIVLPTFTFRWTKSFTL-----ILOALDMNTSY----- 195

QY 134 GLGLMSARLAILGGVEGQPAKGKELVSPCLSPSPFKG-----EWE-----HAEVTVYR 180

Db 196 -----PDAERLIE-----ETSYSVILPSPENKTLDHIGNRARIYR 232

QY 181 VSCQKAAASLHELIC 194

Db 233 VRVQCAVYYNTTC 246

RESULT 15

US-08-346-126-9

; Sequence 9, Application US/08346126

; Patent No. 5849869

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, Spyridon et al.

; TITLE OF INVENTION: Human No. 5849869ch And Delta, Binding Domains

; TITLE OF INVENTION: In Toporythmic Proteins, And Methods Based Thereon

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie &amp; Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/346,126

; FILING DATE:

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/791,923

; FILING DATE: 14-NOV-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Mierock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 7326-007

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212 790-9090

; TELEFAX: 212 8698864/9741

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 9:

; SEQUENCE CHARACTERISTICS:

LENGTH: 293 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-346-126-9

Query Match 4.9%; Score 81; DB 2; Length 293;  
Best Local Similarity 21.6%; Pred. No. 2.9;  
Matches 42; Conservative 21; Mismatches 65; Indels 66; Gaps 6;

QY	14	PLVKAATAGADLLPGEKAKQKLLQKVEVFKASPFVGGDDCVSMNGTDEGVSAWITINFL	73
DB	106	PAELPATITICSPCTHFR-----CLKEYQTTEQASISITGCSFG	147
QY	74	TGSLKTPGGSSVGMLDLGGSTQIAFLPRVEGTLQASPPGYLTALRMFNRTYKLYSYL	133
DB	148	NATTKILGSSSFVLSDPGVGAIVLPFTFRWTKSFLL--ILQALDMYNTSY-----	195
QY	134	GLGMSARLAILGVEGQPAQDGKELVSPCLSPSPFKG-----EWE-----HAEVTVR	180
DB	196	-----PDAERLIE--ETISYGVILLPSPEWKTLDHIGRWARITTR	232
QY	181	VSGQKAAASLHELQ	194
DB	233	VRVQCAVTYYNTTC	246

Search completed: June 8, 2004, 10:11:13  
Job time : 24 secs